Fabre•Kramer Pharmaceuticals, Inc.

Gepirone Hydrochloride Extended-Release Tablets

NDA #21-164

Psychopharmacologic Drugs Advisory Committee Meeting Briefing Materials

December 1, 2015

ADVISORY COMMITTEE BRIEFING MATERIALS: AVAILABLE FOR PUBLIC RELEASE WITHOUT REDACTION

TABLE OF CONTENTS

Contents

I. LIST OF TABLES IN TEXT

II. LIST OF FIGURES IN TEXT

III. LIST OF ABBREVIATIONS

Abbreviations	Definitions
AC	Active control (used in figures and tables only)
AE	Adverse event
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
BMS	Bristol-Myers Squibb
CDER	Center for Drug Evaluation and Research
CGI	Clinical Global Impression
CGI-I	Clinical Global Impression-Improvement
CGI-S	Clinical Global Impression-Severity
CMH	Cochran-Mantel-Haenszel
CSFQ	Changes in Sexual Functioning Questionnaire
CYP450	Cytochrome P450
DALYs	Disability-adjusted life years
DISF-SR	Derogatis Inventory for Sexual Dysfunction – Self-Report
DSM-IV	Diagnostic and Statistical Manual, Fourth Edition
EOT	End-of-trial
ER	Extended-release
ET	End-of-treatment
FDA	Food and Drug Administration
FKP	Fabre-Kramer Pharmaceuticals
HAMD	Hamilton Depression Rating Scale
IR	Immediate Release
ITT	Intention-to-treat
LOCF	Last-observation-carried-forward
MADRS	Montgomery-Åsberg Depression Rating Scale
MAOI	Monoamine oxidase inhibitor
MDD	Major depressive disorder
MDD-AF	Atypical depression
MedDRA	Medical Dictionary for Regulatory Activities
NDA	New Drug Application
OC	Observed Cases
PBO	Placebo (used in figures and tables only)
PC	Placebo control (used in figures and tables only)
PDAC	Psychopharmacologic Advisory Committee
RR	Risk ratio
SD	Standard deviation
SE	Standard error of the mean
SNRI	Serotonin-norepinephrine reuptake inhibitor
SSRI	Selective serotonin re-uptake inhibitor
U.S.A.	United States of America
vs.	Versus

1. EXECUTIVE SUMMARY

Gepirone hydrochloride extended release (gepirone ER) is a new chemical entity under development by Fabre Kramer Pharmaceuticals (FKP) for short-term treatment of major depressive disorder (MDD). We welcome the Psychopharmacologic Advisory Committee's (PDAC's) evaluation of the efficacy and safety data on gepirone ER, as well as your input on key scientific and statistical issues.

On May 1, 2007, FKP re-submitted the NDA for gepirone ER including results from a second positive study [FKGBE007], building on its data from an earlier submission presenting successful results from another positive study [134001]. In this briefing book, we summarize the data and analyses necessary to assess the efficacy of gepirone ER for the treatment of MDD. Specifically, we provide evidence from 2 studies that meet the traditional threshold for FDA approval, as well as detailed supportive information on the safety and effectiveness of the product.

Depression is a major public health problem in the US and around the world that is associated with significant psychiatric and somatic comorbidity, increased mortality, and a leading cause of disability-adjusted life years (DALYs). Despite the availability of several classes of antidepressant agents, many patients do not receive adequate treatment of their depressive symptoms. For example, in one analysis using data from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial, the cumulative remission rate after four sequential acute treatment steps was 67% (Rush 2006). Also, many patients discontinue treatment prematurely due to intolerance of side effects (e.g., sexual dysfunction). Gepirone ER's novel mechanism of action – agonist activity at the 5-HT1_A receptor – provides an alternative to the mechanism of many currently approved antidepressants. While other antidepressants have some agonist activity at the 5-HT1_A receptor (among other actions), this is gepirone ER's sole mechanism of action, making the compound unique among other antidepressant agents.

Despite the fact that the data set supporting the efficacy of gepirone ER is similar to the data set for other, currently approved antidepressant medications, there has been much discussion between FDA and the sponsors as to whether substantial evidence of effectiveness has been demonstrated for gepirone ER. In this Briefing Book, we present the most up-to-date assessment of the totality of the data and include some new analyses performed by leading statisticians to address previous FDA concerns regarding the characterization of certain studies.

There has been considerable discussion with the Agency regarding the characterization of the gepirone ER studies supporting effectiveness. Importantly, there is no disagreement between FKP and the Agency on 8 of the 12 studies. FKP and its advisors have carefully reviewed all of the available data and conclude that of the 12 short-term, randomized, placebo-controlled gepirone ER clinical studies that have been conducted, five of the trials should be considered to be interpretable [134001, 134002, FKGBE007, FKGBE008, and 134023]. These five trials therefore appropriately serve as the basis for an assessment of the efficacy of gepirone ER. Of the five interpretable studies, two clinical trials demonstrated statistically significant findings of gepirone ER's effectiveness on the pre-specified primary endpoint: difference in adjusted mean change from baseline in HAMD-17 score at end of study. The effect size in these studies – as defined by this mean change in HAMD-17 scores – is comparable to the effect size for many of

the currently approved antidepressants, including (Pristiq (desvenlafaxine), Celexa (citalopram), Cymbalta (duloxetine), and Viibyrd (vilazodone)).

Of the twelve short-term studies, we have determined that the remaining seven trials [CN105-078, CN105-083, CN105-052, CN105-053, 134004, 134006, and 134017] should be considered uninterpretable and therefore excluded from an assessment of efficacy.

It is important to note that the clinical development programs for most other currently approved antidepressants - including Viibyrd (vilazodone), Cymbalta (duloxetine), and Celexa (citalopram) - also resulted in the generation of uninterpretable studies (sometimes called failed studies). Additionally, interpretable studies for other currently approved antidepressant development programs typically consist of at least as many negative trials as positive trials.

As stated above, two of the five interpretable studies demonstrate gepirone ER's effectiveness on the primary endpoint (difference in adjusted mean change from baseline in HAMD-17 score at endpoint). The most scientifically sound approach to evaluate the totality of the evidence supporting the efficacy of gepirone ER is to perform a series of meta-analyses on all five interpretable studies. These analyses demonstrated statistically significant benefits of gepirone ER on the following endpoints: (1) average difference in adjusted mean HAMD-17 score; and (2) response rate as defined as a greater than 50% reduction in HAMD-17 score. Taken together with the findings from the two positive studies and the additional supportive data, these meta-analyses provides compelling evidence of the effectiveness of gepirone ER in MDD.

Gepirone ER has an excellent safety profile. Not unexpectedly, based on its serotonergic mechanism of action (albeit unique in its specificity), the body systems most frequently affected are similar to those of other serotonergic agents. AEs reported by at least 5% of all subjects included headache, dizziness, somnolence, nausea, dry mouth, diarrhea, constipation, insomnia, upper respiratory tract infection, nasopharyngitis, and fatigue.

However, gepirone ER's benefits far outweigh its risks. In trials conducted to date, there is no evidence that treatment with gepirone ER causes certain side effects common to other approved serotonergic agents including withdrawal effects, high incidence of somnolence, weight gain, seizure risk, and sexual dysfunction. Having fewer, less severe side effects – including a profile similar to placebo for sexual dysfunction – while providing comparable antidepressant efficacy, gepirone ER may offer clinical advantages over other currently approved antidepressants. In particular, the sexual dysfunction associated with traditional antidepressants – reported to be in the range of 40% (Higgins 2010) – has greatly limited their utility.

Taken as a whole, data from five well-controlled, adequately powered, randomized, interpretable short-term studies provide substantial evidence of effectiveness for gepirone ER in the acute treatment of MDD. The data from this program also provides evidence that gepirone ER is well tolerated and safe for patients with MDD, and possesses a strongly positive benefit risk profile.

2. INTRODUCTION

2.1 Medical Need for New Anti-Depressant Treatments

Depression is a common disorder; in the U.S.A in 2013 an estimated 15.7 million adults aged ≥18 years old had at least one major depressive episode in the previous year, representing 6.7% of all adults.[1] MDD is commonly referred to as depression, and is a mental disorder characterized by mood changes and other symptoms that interfere with a person's ability to work, sleep, study, eat and enjoy once-pleasurable activities. Nearly one-quarter of the population of the U.S.A. is projected to experience MDD during their lifetime (Kessler 2005).

Approximately 50% of those with MDD have severe to very severe clinical manifestations, and over one-half report severe or very severe impairment. The median duration of major depression episodes varies significantly between 3 months and 1 year, depending on severity of the condition (Angst 1995a, Angst 1995b, Furukawa 2003, Solomon 1997, Weissman 1996), and approximately 20% of those with MDD have a duration exceeding 2 years (Spijker 2002). Recovery from a major depressive episode is often incomplete, leading to residual sub-syndromal symptoms as detrimental as the syndrome itself (Israel 2010, Mann 2005). There is no risk-free period of depression, since the risk of developing depression remains relatively constant from adolescence through the lifespan (Kessler 2003).

Suicide, which is often associated with depression, is an important public health problem. In 2013, there were over 40,000 suicides in the U.S.A. (CDC 2013). Furthermore, studies have consistently shown that depression is the most common psychiatric illness found in those who commit suicide (Isacsson 2006). Attempted suicide is also closely associated with psychiatric disorders.

Depression is also associated with significant psychiatric and somatic comorbidity (Evans 2005), as well as increased mortality. Economically, depression is associated with significant costs to the healthcare system. In 2010, the estimated direct costs to health systems was over $95 billion USD (Greenberg 2015). In addition to direct healthcare costs, which are higher among patients with depression, costs include indirect sequelae of depression, such as lost work productivity and disability, which range from $31 to 52 billion annually (Greenberg 2015). Finally, depression is the leading cause of disability-adjusted life years (DALYs) in middle- and high-income countries, and third globally (World Health Organization 2008).

Despite the availability of several classes of antidepressant agents, many patients fail to receive adequate treatment for alleviation of their depressive symptoms (Rush 2006). Remission rates remain low, as highlighted in an analysis of data from the STAR*D trial; the cumulative remission rate after four successive acute treatment steps was 67% (Rush 2006). The limitations of current therapies for MDD have been reviewed at length in the medical literature, including an article published under the auspices of the Center for Drug Evaluation and Research (CDER)

[1] *National Institute of Mental Health - Major Depression Among Adults: http://www.nimh.nih.gov/health/statistics/prevalence/major-depression-among-adults.shtml*

(Laughren 2010). In particular, the article makes the point: "There have been no real "breakthrough" drugs since the SSRIs/SNRIs and the atypical antipsychotics. Most psychiatric new drug approvals in recent years have not been "novel" compounds, rather, active enantiomers of already approved racemic mixtures, active metabolites of parent drugs that have activity very similar to the parent, or other "me-too" drugs (i.e., members of the same class with minor differences)" (Laughren 2010).

Gepirone ER represents an important potential addition to the treatment armamentarium, as it is the only treatment for depression that relies solely on the 5-HT1$_A$ agonism mechanism of action. In addition to its favorable side effect profile, this novel mechanism allows the possibility that patients who have not obtained relief via other antidepressant mechanisms, will find relief with gepirone ER.

2.1.1 Unique Sexual Functioning Profile

Sexual dysfunction contributes to the morbidity of depressive illness, affecting levels of concomitant illness and quality of life, as well as reproductive ability in younger populations. Studies have also shown that patients with a combination of depression and sexual dysfunction are more prone to suicide (Nurnberg and Hensley 2003).

Approximately three of four depressed patients have sexual dysfunction prior to antidepressant treatment (Clayton 2006). It is important that this aspect of depression be treated. Traditionally, however, physicians have not pursued the question of sexual dysfunction in depressed subjects (Harsh 2008). Physicians may not do a thorough evaluation of sexual function at initial intake prior to treatment (Kingsberg 2009). Many physicians may be unaware of how pervasive and serious sexual dysfunction is in the population of depressed subjects. Unfortunately, as clinicians are now becoming aware, currently available antidepressants with selective serotonin re-uptake inhibitor (SSRI) or serotonin-norepinephrine reuptake inhibitor (SNRI) mechanisms further interfere with sexual function (Higgins 2010).

FDA highlighted the disconnect that exists between current antidepressant labeling and actual clinical incidence of treatment-emergent sexual dysfunction in its August 2012 Regulatory Science Forum. Up to 90% of patients with SSRI treatment-emergent sexual dysfunction will discontinue their prescribed medication prematurely (Nurnberg 2008), placing patients at increased risk for recurrence, relapse, chronicity, and mortality (e.g., suicide) (Nurnberg & Hensley 2003). Studies also suggest that sexual dysfunction caused by SSRIs may continue after the drug is discontinued and have a permanent impact (Csoka 2008, Kennedy and Rizvi 2009).

Thus, there is an unmet medical need for an antidepressant without sexual dysfunction as a side effect and, ideally, with the ability to improve sexual function while also alleviating depression. None of the three recently approved antidepressants, vilazodone, levomilnacipran, and vortioxetine has addressed this issue, as all have SSRI or SNRI pharmacology as their primary mechanism of action, and all have higher incidence of sexual side effects than placebo.

Gepirone ER represents an important additional option to the physician's armamentarium. Consistent with its mechanism of action (i.e. agonist activity at the 5HT$_{1A}$ receptor, as opposed

to broad-based serotonin reuptake inhibition), sexual functioning data collected in gepirone ER studies indicates gepirone ER does not induce treatment-emergent sexual dysfunction. These findings, coupled with a better sexual dysfunction side effect profile when compared with the SSRIs, can be expected to improve treatment compliance of patients with MDD, resulting in improved treatment outcomes with gepirone ER relative to those achieved with currently available antidepressants (Fabre 2011, Fabre and Smith 2012).

2.2 Clinical Study Challenges in Antidepressant Development

MDD is a heterogeneous illness, and recent clinical trials demonstrate the continued challenges developers of new treatments face in generating reliable data for demonstrating effectiveness.

Recent estimates suggest that approximately 50% of the trials in Phase III clinical development programs for drugs receiving FDA approval for antidepressant use in MDD failed to demonstrate superiority over placebo (Khin 2011). The reasons for the high failure rates remain somewhat elusive, but the sources of variability in outcomes for clinical trials of psychiatric medications are thought to fall into three general areas: trial design characteristics, study participant characteristics, and quality of study conduct (Dunlop 2015). Taken together, it is clear that developing drugs for MDD is a challenge, but there remains a large unmet medical need for novel antidepressants, in spite of a number of recent antidepressant approvals by the FDA.

The dataset for gepirone ER is similar to the datasets for other recently approved antidepressants. For example, vilazodone was approved by the FDA in 2011 for the treatment of MDD, despite a challenging clinical development program. The clinical development program in support of efficacy included two pivotal phase III trials, and five additional trials; including three studies that incorporated both placebo and active comparator, and two studies that included placebo only. Notably, all five trials failed to demonstrate superiority to placebo on the primary endpoint. Since none of the active comparators demonstrated superior efficacy compared with placebo, these three-arm trials were recognized as failed studies. Furthermore, the two studies with placebo-only were recognized as "negative" studies (Laughren 2011, Wang 2015).

As a point of comparison, FKP reviewed the three-arm short-term comparator trials for three recently FDA approved antidepressants (Table 1). From this analysis, it is clear that the geprione ER dataset is not unique. Indeed, for two of the three approved NCE antidepressants (i.e. citalopram and vilazodone), 100% of the trials were deemed failed since they lacked assay sensitivity. Furthermore, when all three drugs are considered, > 90% of the trials (13/14) were considered failed. Importantly, data from these trials were not included in the efficacy evaluations of these drugs.

Table 1: Three Recent FDA Approved NCE Antidepressants - Characterization of All Short-Term 3 Arm Comparator Trials

Investigational Drug	Year Approved	Short-Term 3 Arm Comparator Studies	Failed*	Positive**	Percentage Failed
citalopram	1998	5	5	0	100
duloxetine	2004	6	5	1	83
vilazodone	2011	3	3	0	100
TOTALS		**14**	**13**	**1**	**94**
gepirone ER		5	5	0	100
ADJUSTED TOTAL		19	18	1	96

Source: **FKP - Data presented at FDA meeting (November 2011)**

As shown above, during the clinical development of many NCE antidepressants, virtually all active control three-arm comparator studies fail. A failed three-arm trial is one that is uninterpretable because the comparator fails to separate from placebo on the primary endpoint(s) and the study therefore lacks assay sensitivity (FDA 1998).

Increasingly, another reason many antidepressant clinical trials are unsuccessful in demonstrating efficacy for the test drug (whether or not there is an active control arm) is high placebo response rates. Khan, et al, showed that in studies of approved antidepressants, only 21% of studies where the placebo response was > 30% demonstrated efficacy for the test drug versus 74% of studies with placebo response rates < 30% (Khan 2003). There were no positive studies among the approved antidepressants in the Khan database when the placebo response was >38%. Those included studies of 5 different approved drugs.

While grappling with complex datasets can present a challenge, it is important that results from uninterpretable studies not cloud the evidence provided by the interpretable studies. To reduce subjective bias, these data should be handled in a scientifically appropriate manner, since this type of evaluation becomes crucial when evaluating the totality of evidence in support of the effectiveness of a novel antidepressant.

2.3 Gepirone ER Formulation Characteristics

Gepirone ER is 4, 4-dimethyl-1-[4-[4-(2-pyrimidinyl) 1-piperazinyl] butyl]-2, 6-piperidinedione hydrochloride. As a member of the azapirone class of chemicals, gepirone ER is a full agonist at the pre-synaptic 5-HT_{1A} autoreceptor and a partial agonist at the post-synaptic 5-HT_{1A} receptor.

Originally, gepirone ER was developed as an immediate-release (IR) formulation. However, during early-phase comparisons between the gepirone IR and ER formulations, pharmacokinetic differences (i.e., reduced frequency of administration and more consistent plasma levels (Timmer 2003)) were observed for the ER formulation that may contribute to therapeutic advantages for the patient. These pharmacokinetic differences may allow improved adherence to therapeutic regimens, an improved safety profile and improved efficacy through the achievement of higher doses (Rush 2004). Therefore, the development of gepirone IR was discontinued in favor of the reformulated ER.

Gepirone ER uses a hydroxypropyl methylcellulose formulation to mediate extended release of the active ingredient, which facilitates absorption over 16 hours. This extended release mechanism allows for once daily administration. This confers an improvement in patient adherence to antidepressant pharmacotherapy - a major advantage for the use of gepirone ER. For example, in a pooled analysis of depression studies, it was found that the odds of patients on a once-daily regimen being adherent was more than twice the odds of being adherent for patients on twice-daily dosing (Medic 2013).

The elimination half-life of gepirone ER is 6 to 11 hours, and, when taken in the presence of food, area under the curve (AUC) increases by up to 37%. Gepirone ER undergoes extensive first-pass metabolism, with formation of two major pharmacologically active metabolites, 3-OH-gepirone, and 1-pyrimidinyl piperazine (1-PP). 3-OH-gepirone is the active metabolite as an agonist at the 5-HT_{1A} receptor. Gepirone ER is metabolized mostly by cytochrome P450 (CYP450) 3A4, with minor metabolism occurring at CYP450 2D6. Gepirone ER is not known to be an inducer or inhibitor of major CYP450 metabolic pathways.

2.4 The Unique Mechanism of Action of Gepirone ER

Gepirone ER is a full agonist at the pre-synaptic 5-HT_{1A} auto-receptor and a partial agonist at the post-synaptic 5-HT_{1A} receptor (Fitton 1994; Amsterdam 1992). This unique pharmacological mechanism offers a new approach to the modulation of serotonergic neurotransmission in the treatment of depression. Gepirone ER does not inhibit the reuptake of serotonin, or noradrenaline, nor does it inhibit monoamine oxidase. Gepirone ER has only weak affinity for adrenergic and histaminergic receptors.

Gepirone ER is structurally unrelated to SSRIs, tricyclics, tetracyclics, or monoamine oxidase inhibitors. The class of azapirone drugs is utilized for psychiatric disorders and they exert their effects by being partial agonists of 5-HT_{1A} receptors (Stahl 1993). Treatment with gepirone ER desensitizes presynaptic 5-HT_{1A} receptors, which decreases serotonin autoregulatory inhibition and enhances activation of postsynaptic 5-HT_{1A} receptors (Fitton 1994). As a partial agonist, gepirone ER acts as an agonist when endogenous serotonin is not present and as an antagonist when endogenous serotonin is present (Stahl 1993). Overall, gepirone ER increases serotonin production when insufficient amounts are present, and decreases serotonin production when excess amounts are present (Stahl 1993).

Gepirone ER has weak to no affinity for other receptor sites, and unlike other azapirones, gepirone ER does not possess clinically relevant anti-dopaminergic affinity for the dopamine D2 receptor (Fitton 1994). That is, gepirone ER does not elevate levels of dopamine, a neurotransmitter that when present at high levels is associated with addiction (Sulzer 2011). This lack of affinity for dopamine receptors indicates that gepirone ER has a very low abuse potential (Fabre 2012).

Finally, based upon its unique mechanism of action at 5-HT_{1A}, gepirone ER also might be expected to increase libido and as well as possessing anxiolytic effects (Fabre 2012). These represent significant features since loss of libido and increased anxiety are common symptoms of depression and common side effects of many existing antidepressants. For example, SSRIs are known to functionally increase serotonin at all serotonin receptors. However, increased activity

at 5-HT2A receptors is particularly damaging to sexual function (Pfaus 2009, Bishop 2006). On the other hand, gepirone ER does not affect other serotonin receptors, and improves sexual function, likely through its activity at the 5-HT1$_A$ receptor.

The unique mechanism of action of gepirone ER also contributes to its favorable safety profile. Overall, more than 6,000 subjects have received gepirone (IR and ER formulations) and the only significant side effects have been transitory light-headedness and nausea, which are common side effects among approved antidepressants. Importantly, gepirone ER does not have withdrawal effects, cardiovascular issues, or seizure risk, and does not cause weight gain; all problematic side effects of currently marketed antidepressants.

2.5 Gepirone ER Clinical Development and Regulatory History

A development plan for gepirone ER was initiated by Bristol-Myers Squibb (BMS) in 1991 intended to address tolerability and dosage issues observed during clinical studies with the IR formulation. The specific objectives for development of gepirone ER were to reduce the incidence of adverse events and to improve adherence to the medication regimen by decreasing the frequency of administration. BMS discontinued all gepirone ER development in 1992 for business reasons. Four ongoing Phase III studies evaluating its use in the treatment of MDD were prematurely terminated. These were the initial Phase III studies of gepirone ER in MDD, and the effective dose range had not been determined.

FKP acquired the rights to gepirone ER in 1993, and completed several Phase I studies in special populations. In 1998, FKP licensed gepirone ER to Organon Pharmaceuticals for further development and commercialization. Organon began a phase III development program with two registrational studies that completed in late 2000 [134001 and 134002]. One of these studies (134001) met the pre-specified primary endpoint, and based on prior discussions with FDA, Organon expected this study, combined with earlier IR studies, would be sufficient for approval. Organon initiated a group of additional trials designed to evaluate gepirone ER in atypical depression and to collect comparative data [134004, 134006, and 134017]. However, Organon was unsuccessful in obtaining FDA approval, terminated its development of gepirone ER in 2004 and returned licensing rights to FKP in 2005.

In late 2003, FKP, with Organon's consent, initiated two additional double-blind, placebo-controlled registrational studies of gepirone ER in MDD [FKGBE007 and FKGBE008], one of which met its primary efficacy endpoint (FKGBE007). After completing these trials and reacquiring the rights from Organon, FKP agreed with FDA to submit an ER only NDA for gepirone incorporating the two successful ER studies [134001 and FKGBE007] In addition, FKP analyzed and reported on the beneficial effects of gepirone ER on sexual function. FKP submitted the NDA in May 2007 and received a not approvable letter in November 2007.

During this period, it has also been discussed with the FDA how the totality of the evidence should be evaluated, which has culminated in this Advisory Committee review. Briefly, in May 2011, FKP filed a request for reconsideration and included a statistical report by an independent statistical consultant. In the report, the consultant evaluated each of the short-term clinical investigations on gepirone ER for the treatment of MDD. The consultant concluded that only 5 clinical trials should be considered as interpretable in an evaluation of the efficacy of gepirone

ER. FKP and FDA met in November 2011 to discuss the reconsideration. In February 2012, at request of FDA, FKP submitted an NDA amendment that contained additional support. After further discussions, FDA requested FKP submit another amendment to the NDA incorporating information on all short term trials as well as detailed data on gepirone ER's sexual functioning characteristics. FKP submitted this NDA amendment in December, 2012.

FDA issued a General Advice letter to FKP in April 2014 concluding its reconsideration, continuing to question the effectiveness of gepirone ER and reiterating the clinical deficiencies from the 2007 non-approvable action.

In January 2015, FKP's request for formal dispute resolution by the Office of New Drugs (OND) appealing the November 2, 2007 Not Approvable letter and the April 2014, General Advice Letter was accepted. A meeting was held with Dr. John Jenkins, Director OND, on February 23, 2015, after which he provided an interim response indicating he had asked for additional review of the data within FDA. Thereafter, in a June 1, 2015 letter Dr. Jenkins indicated the need for input from the PDAC before he could reach a decision on the appeal.

Given that MDD is a difficult condition to study, and most approved antidepressants have had multiple negative and/or failed studies in their Phase III development programs, FKP looks forward to an open discussion with the PDAC on how to best evaluate the totality of evidence supporting the efficacy of gepirone ER in the treatment of MDD.

3. EFFICACY

3.1 Overview of Clinical Studies

The clinical development program of gepirone ER for the treatment of MDD included the 12 short-term studies summarized in Table 2 below.

Table 2: Gepirone ER – Summary of Short-term Clinical Studies

Study Number	Number of Subjects (ITT)		Active Control	LS mean diff HAMD-17 p-value (gep-ER vs. Pbo)	Protocol defined endpoint† p-value		Summary
	gep ER	Pbo			gep-ER vs. placebo	Active vs. placebo	
134001	101	101	None	-2.47; p=0.013	-2.47; p=0.013	--	Significant treatment effects for primary and secondary efficacy variables (p<0.05).
FKGBE007	116	122	None	-2.45; p=0.018	-2.26; p=0.032	--	Significant treatment effects for primary and secondary efficacy variables (p<0.05).
134002	102	103	None	-0.71; p=0.42	-0.67; p=0.446	--	Positive trends for all vars; mixed models show p<0.05 for mHAMD-17, Bech-6, HAMD item 1, and MADRS
FKGBE008	96	99	None	-1.38; p=0.20	-1.5; p=0.159	--	Positive trends for all vars; p<0.05 for HAMD-17 (Wks 2, 3, 6) and MADRS (Wks 2, 3, 4, 6)
134023	123	123	None	0.13; p=0.90	0.13; p=0.898	--	No trends or significance for any variables.
CN105-078*	88	47	None	-1.0; p=0.36	-0.9; p=0.451	--	Terminated early, 62% power; positive trends for high dose.
CN105-083*	73	39	None	-0.49; p=0.75	-0.5; p=0.742	--	Terminated early, 53% power; positive trends for high dose.
CN105-052*	35	37	fluoxetine	-0.69; p=0.74	-0.66; p=0.757	-0.5; p=0.798	Terminated early, 43% power; no significant treatment effects; comparator failed on primary endpoint (no assay sensitivity).
CN105-053*	56	56	imipramine	-2.0; p=0.19	-0.70; p=0.687	-2.50; p=0.144	Terminated early, 63% power; only 1 site (of 2) completed enrollment and showed positive efficacy for gepirone ER and imipramine.
134004	124	130	fluoxetine	1.04; p=0.18	0.87; p=0.416 HAMD-25	-1.03; p=0.325 HAMD-25	MDD-AF; comparator failed on primary endpoint (no assay sensitivity); high placebo response (42%)
134006	140	143	paroxetine	0.22; p=0.76	0.06; p=0.953 HAMD-25	-1.58; p=0.178 HAMD-25	MDD-AF, comparator failed on primary endpoint (no assay sensitivity); high placebo response (46%)
134017	159	159	fluoxetine	0.65; p=0.39	0.50; p=0.650 MADRS	-1.15; p=0.299 MADRS	Comparator failed on primary endpoint (no assay sensitivity); high placebo response (53%)

†Based on primary efficacy endpoint and method of analysis pre-specified in the protocol; the primary endpoint is HAMD-17 unless otherwise stated.
*Terminated early due to business/administrative reasons by Bristol-Myers Squibb.
MDD-AF = Atypical Depression. Placebo response="much" or "very much" improved on CGI change score.

3.2 Selection of Studies for Assessment of Efficacy

The 12 available clinical trials vary widely in terms of their quality and relevance to the determination of efficacy of gepirone ER.[2] While it is desirable to draw upon all data that might shed light on this important issue, it is evident that studies of antidepressants typically entail serious methodological obstacles that often impair study validity. Consequently, it is ultimately a matter of careful scientifically sound evaluation to determine which studies met scientific and statistical criteria necessary to generate interpretable data that ought to contribute to decision-making.

Given the statistical issues that have become central to this review, FKP, after consulting with a group of statistical advisors, had a new independent assessment of the clinical data conducted. This assessment, summarized below, produced conclusions consistent with a previous independent analysis conducted in 2011 by another statistical consultant (The 2011 report is provided as Appendix 4).

The Phase III development program of gepirone ER for the treatment of MDD was complicated by the involvement of multiple sponsors and associated interruptions in the time-span of the program. There are a number of factors that that have come into play that affect the ability to draw clear and valid inferences. Importantly, as will be explained, virtually all of these factors have tended to operate in the direction of attenuating the estimate and/or statistical significance of gepirone ER's effect relative to placebo or an active comparator. Nonetheless, the assessment concluded that there are five studies that yielded interpretable effect estimates. Two of these were successful in meeting the primary efficacy endpoint.

- 134001
- FKGBE007
- 134002
- FKGBE008
- 134023

The remaining seven studies were assessed to have major deficiencies that undermine the interpretability of their results, for reasons specified below and further elaborated in Appendix 2:

- CN105-078
- CN105-083
- CN105-052
- CN105-053
- 134004
- 134006
- 134017

[2] FKP and FDA are in agreement with regard to the interpretability of 8 of the 12 studies. The open issues between FDA and FKP seem to involve the remaining 4 short-term active-controlled studies: ORG134004, ORG134006, ORG134017, and CN105-053. As described in this briefing book, FKP considers them failed studies because neither gepirone ER nor the active control beat placebo on the protocol-specified endpoint using the protocol-specified analysis.

Table 3 contains a summary of the salient data pertaining to study quality. The purpose of this table is to lay out the potential threats to validity of all 12 studies, and enable a determination of those that are interpretable. More specifically, an interpretable study is one that provides unbiased valid estimate of the efficacy of gepirone ER relative to placebo when administered in adequate therapeutic doses to an appropriate population of MDD patients. Based on this criterion, 7 studies were deemed to have not met this standard.

Table 3: Summary of Gepirone ER Short-Term Studies – Comparison of Factors Influencing Study Validity

Study	Power	Baseline HAMD-17	Mean Dose	% Dropout Gepirone ER	% Dropout Placebo	Assay Sens.	%HAMD Pbo Resp	%CGI Pbo Resp
134001 *+	80%	22.7	61.1	27.5	23.6	N/A	29.7	35.6
134002 *+	80%	24.0	57.9	31.8	28.7	N/A	38.5	44.7
134023 *+	85%	22.9	61.3	26.0	21.3	N/A	35.1	39.0
FKGBE007 *+	85%	23.9	58.2	21.8	17.8	N/A	32.9	34.7
FKGBE008 *+	85%	24.2	60.0	24.0	21.5	N/A	34.9	37.8
CN105-078 +	62%	22.3	40.5	42.1	30.6	N/A	28.4	38.3
CN105-083 +	53%	23.9	43.9	38.1	34.1	N/A	37.1	43.6
CN105-052 +	43%	25.2	43.4	38.9	50.0	No	41.7	56.8
CN105-053	63%	23.9	46.4	41.4	60.7	?	44.8	56.3
134004	80%	19.6	67.1	36.3	21.3	No	38.8	42.3
134006	85%	19.0	55.3	31.3	24.3	No	42.0	46.9
134017	85%	23.3	58.9	31.5	21.3	No	46.6	52.8

*Interpretable Study
+ Studies for which FKP and FDA are in agreement with regard to interpretability

The interpretable clinical studies are summarized broadly in the following sections and in more detail in Appendix 1. The uninterpretable studies are also summarized broadly in Section 3.4 and in more detail in Appendix 2.

In addition, there has been one long-term, placebo-controlled relapse prevention study (28709) for gepirone ER. More details for study 28709 are provided in Section 3.7 and Appendix 3.

3.3 Summary of Individual Interpretable Gepirone ER Studies

In the following sections, we provide a summary overview of the gepirone ER studies that the independent assessment considered to be interpretable.

3.3.1 Study 134001

This was a 5-center, randomized, double-blind, placebo-controlled study evaluating gepirone ER 20-80 mg/day over an 8-week treatment period in subjects with MDD. This study employed doses in the appropriate therapeutic range for gepirone ER. Study 134001 was adequately designed and executed, and the primary efficacy parameter was the change from baseline in HAMD-17 total score based on the LOCF analysis of the ITT population. The study had one forced-titration step from 20-40 mg/day at Day 4, and a flexible-dose design thereafter. The mean dose (± SD) of gepirone ER was 61.05 (±12.02) mg/day.

The treatment effect was statistically significant favoring gepirone ER over placebo for the primary efficacy variable, and this positive result was supported by nearly all secondary efficacy variables.

Study 134001 provides strong evidence of the therapeutic effectiveness of gepirone ER for the treatment of subjects with MDD, and its effect size [2.5 points on HAMD-17] is comparable to that of other approved antidepressants.

3.3.2 Study 134002

This was a 5-center, randomized, double-blind, placebo-controlled study of gepirone ER 20-80 mg/day over an 8-week treatment period in subjects with MDD. This study employed doses in the appropriate therapeutic range for gepirone ER. The study had a flexible-dose design; the minimum final dose of gepirone ER was 40 mg/day. The mean dose (± SD) of gepirone ER was 57.90 (± 13.03) mg/day. The final prescribed dose was 60 and 80 mg/day in 23.4% and 58.9% of subjects, respectively.

While results for the primary efficacy variable were not statistically significant, secondary efficacy variables of HAMD25 responders and HAMD-Item 1 CFB were statistically significant, there were positive trends for several other secondary endpoints and significant treatment effects were obtained using a post-hoc repeated measures (mixed models) analysis, providing supportive evidence that gepirone ER has anti-depressant activity.

3.3.3 Study 134023

This was a 12-center, randomized, double-blind, placebo-controlled, flexible dose trial evaluating gepirone ER in subjects with MDD during an 8-week treatment period. The mean dose (± SD) of gepirone ER was 61.3 (±13.66) mg/day, with 69.3% of subjects reaching a final dose of 80 mg/day.

No statistically significant treatment effects were detected for gepirone ER based on the primary or secondary efficacy variables.

While adequately designed, 134023 is a negative study.

3.3.4 Study FKGBE007

This study was a 9-center, randomized, double-blind, placebo-controlled study, which employed doses in the appropriate therapeutic range for gepirone ER over an 8-week treatment period in subjects with MDD. The study had one forced dose titration step from 20 mg/day to 40 mg/day between Days 4 and 7, and a flexible dose design thereafter. The mean dose (± SD) of gepirone ER was 58.2 (±13.95) mg/day.

The treatment effect was statistically significant favoring gepirone ER over placebo for the primary efficacy variable, and this positive result was supported by nearly all secondary efficacy variables.

Study FKGBE007 provides strong evidence of the therapeutic effectiveness of gepirone ER for the treatment of subjects with MDD, and its effect size [2.5 points on HAMD-17] is comparable to that of other approved antidepressants.

3.3.5 Study FKGBE008

This was an 8-center randomized, double-blind, placebo-controlled, flexible dose study evaluating gepirone ER 20-80 mg/day over an 8-week treatment period in subjects with MDD. The mean dose of gepirone ER was 60.0±13.1 mg/day. By the final visit, 86.9% of subjects were at a dose of 60-80 mg/day.

While this study failed to demonstrate a statistically significant treatment effect for the protocol-defined primary endpoint (change from baseline in HAMD-17 scores at Week 8), trends in mean values directionally favored gepirone ER over placebo at each visit, with significant differences detected at Week 2 and Week 6. By Week 8, the mean change from baseline scores were not significantly different.

These data suggest that gepirone ER has a beneficial impact on symptoms of depression.

3.3.6 Summary of Efficacy Data from Interpretable Gepirone ER Studies

Of the 5 interpretable studies, two clinical trials demonstrate gepirone ER's effectiveness on the primary endpoint: average difference in adjusted mean HAMD-17 score. In addition, 2 of the remaining 3 studies support gepirone ER's effectiveness on various secondary endpoints, including responder rate, defined as greater than 50% reduction in

HAMD-17 score. Therefore, 2 of the 5 interpretable studies provide well-demonstrated evidence of effectiveness and 4 of the 5 interpretable studies provide either well-demonstrated or supportive evidence of effectiveness.

3.4 Summary of Individual Uninterpretable Gepirone ER Studies

In the following sections, we provide a summary overview of the gepirone ER studies that the independent assessment considered to be uninterpretable. The salient features are also summarized in Table 4.

3.4.1 Studies CN105-078 and CN105-083

These studies had similar designs and suffer from similar deficiencies. Both were terminated prematurely by BMS well in advance to reaching planned enrollment levels, resulting in diminished statistical power. Low statistical power would not by itself invalidate the effect estimate or make it uninterpretable. However, it would certainly reduce the chances of obtaining statistical significance at the conventional 0.05 level. Thus, these studies can certainly be considered failed as definitive evaluations of efficacy, although possibly could be deemed relevant to a broader interpretation of the total evidence. However, there is a serious problem that undermines even this limited utility. The level of dosing was clearly inadequate for a large proportion of patients, as evidenced by an average dose close to the 40 mg therapeutic minimum.

3.4.2 Studies CN105-052 and CN105-053

These two studies had essentially the same problems of low power and inadequate dosing as CN105-78 and CN105-83, and these studies can be considered failed as definitive evaluations of efficacy. Both were terminated prematurely by BMS well in advance to reaching planned enrollment levels. In addition, the data related to dropouts and response rates in the placebo arms appear anomalous. The overall dropout rates in the two studies are by far the highest among the 12 studies, and the placebo dropout rates of 50.0% (CN105-052) and 60.7% (CN105-053) suggest a systemic problem. Notably, data from the fully enrolled site in study CN105-053 shows assay sensitivity on the primary efficacy parameter, with significant positive effects for both gepirone ER and the active control (imipramine); the smaller site (only 15-16 subjects per group) received lower doses of study drug and shows no positive effects for either treatment. The largest site supports the efficacy of gepirone ER. The detailed study descriptions in Appendix 2 shed more light on the unusual circumstances that gave rise to these results. Another anomaly is indicated by the extremely high CGI placebo response rates of 56.3% (CN105-052) and 56.8% (CN105-053). Again, more detail about these studies can be found in Appendix 2.

3.4.3 Studies 134004 and 134006

These two studies had similar designs and suffer from the same major limitation. There is clear evidence that the special population chosen was inappropriate for comparison with the other studies. Specifically, the studies were designed to test gepirone ER in patients with MDD with atypical features. All of the other short term studies in the NDA contained few if any such patients. Moreover, the mean baseline HAMD-17 levels in these studies, 19.6 (134004) and 19.0 (134006), were substantially lower than for the other studies as there was no minimum baseline severity criteria for enrollment in either study. Over 50% of the subjects in each study had baseline HAMD-17 below the minimum of all the other studies. Furthermore, in both of the studies the active control failed to separate from placebo on the primary endpoint, indicating the lack of evidence of assay sensitivity. For the above reasons, these studies can be considered failed as definitive evaluations of efficacy, apart from the issue of having enrolled inappropriate patient populations.

3.4.4 Study 134017

This study had lack of assay sensitivity, as there was no significant difference between the active comparator (fluoxetine) and placebo for the primary endpoint, which in this case was mean change in adjusted MADRS score change from baseline. It is therefore a failed study and its data uninterpretable. The extremely high placebo response rate (46.6% on HAMD and 52.3% on CGI) may provide an explanation for the failure of both gepirone ER and fluoxetine to separate significantly from placebo. Although a post hoc analysis by FDA produced a p-value of 0.042 in favor of fluoxetine vs. gepirone ER, it corresponds to opposite signs of the non-significant comparisons of the active treatments to placebo, and so could be spurious in this sense as well as with respect to unplanned multiplicity of comparisons. In addition, this type of comparative efficacy is not a valid approach since showing a difference between two effective drugs is notoriously challenging and usually requires a very large study.

3.4.5 Summary of Uninterpretable Gepirone ER Studies

Of the twelve short-term gepirone ER clinical studies that have been conducted, seven of the trials were assessed to be uninterpretable due to a series of statistical and methodological principles and therefore not appropriate for inclusion in the dataset to evaluate the effectiveness of gepirone ER. The primary reasons for unintpretability of these studies is presented in Table 4.

Table 4: Primary Reasons for Characterizing Seven Studies as Uninterpretable

Study	Primary Reasons for Uninterpretable/Failed Characterization
CN105-078	Study was terminated prematurely when BMS discontinued development of gepirone ER. Thus, the study had reduced power (62% power) and sample size. Also, inadequate mean dose.
CN105-083	Study terminated prematurely when BMS discontinued development of gepirone ER. Thus, the study had reduced power (53% power) and sample size. Also, inadequate mean dose.
CN105-052	Study was terminated prematurely by BMS for business reasons. Thus, the study had reduced power (43% power) and sample size due to early study termination. The active control (fluoxetine) was indistinguishable from placebo based on all efficacy variables. Moreover, the dose of gepirone ER used in this study was low relative to its therapeutic dose range (60-80 mg/day).
CN105-053	Terminated prematurely and the 2 participating study sites show contradictory efficacy results. Ignoring this interaction, the pooled data from the study is a failed trial because it lacks assay sensitivity. Data from the fully enrolled site shows assay sensitivity on the primary efficacy parameter, with significant positive effects for both gepirone ER and the active control (imipramine); the smaller site (only 15-16 subjects per group) received lower doses of study drug and shows no positive effects for either treatment. The largest site supports the efficacy of gepirone ER.
134004	The study is a failed trial due to lack of assay sensitivity, that is, active control failed to separate from placebo on the primary endpoint. Inappropriate patient population was enrolled (MDD with atypical features). Mean baseline HAMD-17 level was 19.6, and substantially lower than for the other short-term studies.
134006	The study is a failed trial due to lack of assay sensitivity, that is, active control did notseparate from placebo on the primary endpoint. Inappropriate patient population was enrolled (MDD with atypical features). Mean baseline HAMD-17 level was 19.0, and substantially lower than for the other short-term studies.
134017	This was a poorly executed study, with highly inconsistent results among sites. The study is a failed trial due to lack of assay sensitivity, that is, active control did not separate from placebo on the primary endpoint.

3.5 Impact of Dropouts on Demonstration of Efficacy

FKP and its advisors believe it is important to highlight the impact of study dropouts as it pertains to all of the twelve available studies. From Table 3, it is clear that generally, the dropout rate for patients receiving gepirone ER tended to be higher than for those receiving placebo. Such an imbalance could be unfavorable for gepirone ER for two reasons. First, analyses of placebo-controlled studies that employ an ITT/LOCF approach can underestimate the true effect. This can occur because an effective treatment is received for only some fraction of the planned full study duration. However, in the studies of gepirone ER, this problem was exacerbated, because of adaptive titrated dosing required to attain a potentially effective dose. Consequently, a dropout in the gepirone ER arm would tend not only to receive a shorter duration of active treatment, but also a lower average dose, than if the dropout had completed the study.

It is noteworthy that the five studies deemed interpretable all have dropout rates that are relatively low and only slightly higher for the gepirone ER patients than for the corresponding placebo patients. In contrast, the 7 uninterpretable studies have dropout rates that are higher overall and/or display larger between-group differences for dropout rates.

3.6 Meta-Analysis of Gepirone ER Studies

As explained above, there is a strong rationale for focusing on the five interpretable studies. These five studies all provide valid evidence regarding the efficacy of gepirone ER in the treatment of MDD. As with all antidepressant programs the results vary across the five studies in terms of effect size and statistical strength.

Therefore, FKP sought guidance from the following independent statistical advisors to design and perform a series of meta-analyses of these five interpretable studies.

- Dr. Lee-Jen Wei, Ph.D.
  Professor of Biostatistics
  T.H. Chan School of Public
  Health Harvard University

- Herb Weisberg, Ph.D.
  President, Correlation Research, Inc.

The results of these meta-analyses are set forth below in section 3.6.1.

3.6.1 Methodology

The data from the five interpretable studies were pooled and evaluated for the following endpoints:

1) Adjusted Mean Change in HAMD-17 from Baseline to End of Treatment

2) HAMD-17 Responder Analysis

Two models were used for each meta-analysis: a random effects model and fixed effect model. These two models make different assumptions about the nature of the studies, and these assumptions lead to different definitions for the combined effect, and different mechanisms for assigning weights to the studies. In general, the random effects model approach is more robust than its fixed effect counterpart. On the other hand, if the fixed effect model assumption is plausible, the results from the fixed effect is more efficient than the random effects procedure.

Random effects models may be used when the fixed effect assumption is considered to be too strict. These models instead assume that treatment effects may differ from study to study according to a specific "prior" distribution such as a normal distribution (DerSimonian and Laird 1986) and instead estimate the "average" treatment difference. In general, random effects models produce more conservative (i.e. wider) confidence intervals than fixed effect models.

Fixed effect models assume that a specific magnitude of treatment effect is present in all studies under consideration. In this framework, the overall treatment effect estimate represents a weighted average of study-level treatment effect estimates, where the weights are proportional to the inverse of the variance of the study-level effect estimate.

Two sets of responder analyses based on HAMD-17 were conducted with the intention of capturing the magnitude of effect that can be considered clinically relevant.

The first responder analysis used the definition of a HAMD-17 responder that was specified in the original statistical study plans; responders were defined as subjects who showed at least a 50% reduction in baseline HAMD-17 at any post-baseline visit. The responder rates are based on all randomized subjects. The intention of the responder analysis is to capture the magnitude of an effect that can be considered clinically relevant. Sensitivity analyses were then conducted that defined treatment response as a $\geq$ 50% reduction at the last scheduled study visit. This sensitivity analysis made the pessimistic assumption that all drop-outs were non-responders.

3.6.2 Results

3.6.2.1 *Mean Change in HAMD-17 from Baseline to End of Study*

Figure 1 and Figure 2 below present the results of the meta-analyses for the primary efficacy parameter, mean change in HAMD-17 from baseline to the end of the short-term, double-blind treatment period.

The treatment differences between gepirone ER and placebo for the adjusted mean change in HAMD-17 from baseline to the end of the treatment period were statistically significant in favor of gepirone ER for both the random effects [-1.34, 95% CI (-2.33, -0.35)] and fixed effect models [-1.32, 95% CI (-2.18, -0.47)], (p=0.008 and 0.002, respectively).

Figure 1: Mean Change in HAMD-17 from Baseline to End of Treatment (Continuous Outcome Variable) Using Random Effects Model

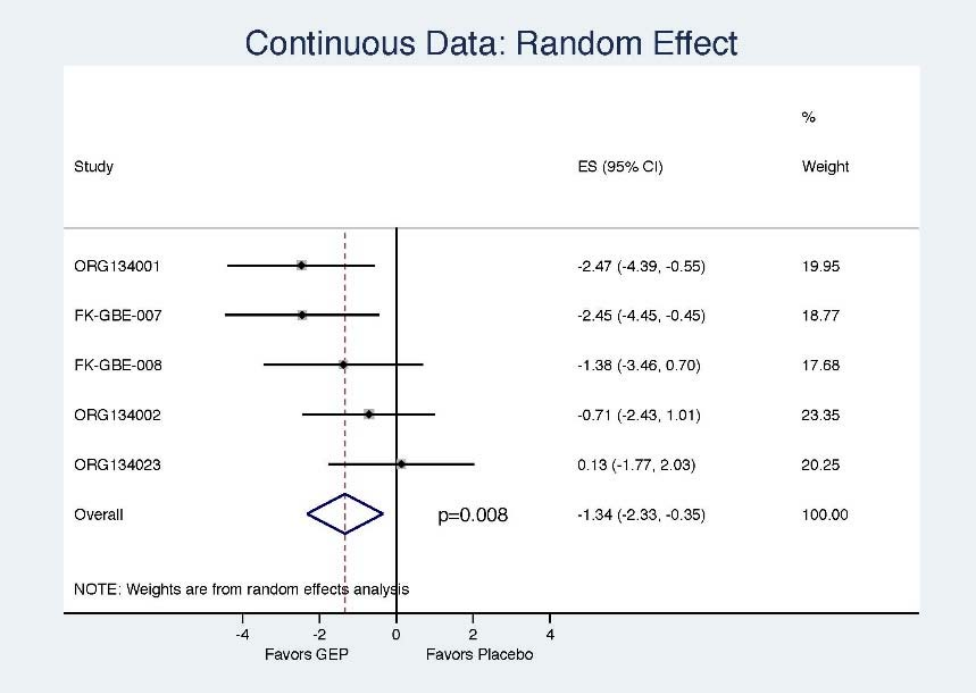
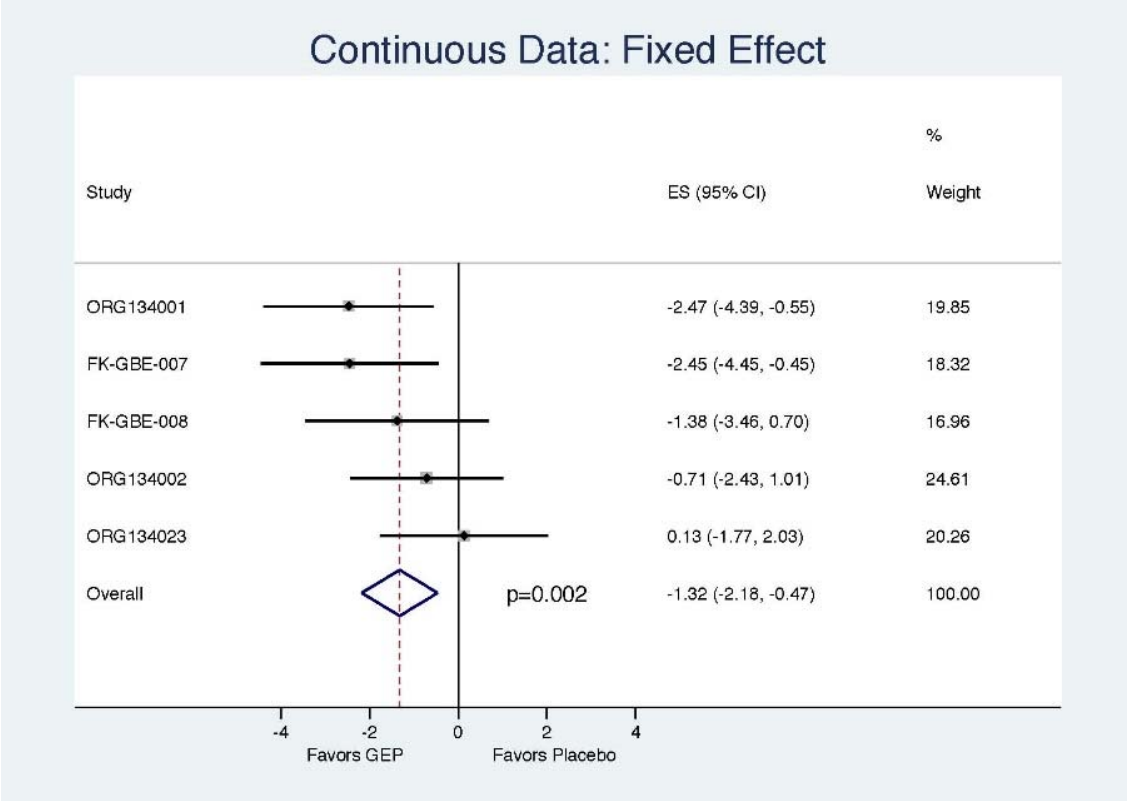


Figure 2: Mean Change in HAMD-17 from Baseline to End of Treatment (Continuous Outcome Variable) Using Fixed Effect Model



3.6.2.2 *HAMD-17 Responder Analysis*

Figure 3 and Figure 4 present results of the meta-analyses for the percent of HAMD-17 responders at any time post-baseline, with the treatment effect quantified using the odds ratio (OR) for treatment response with gepirone ER relative to placebo.

Meta-analysis of the five studies meeting the eligibility requirements for the meta-analysis showed a statistically significant difference in favor of gepirone ER for both the random effects and fixed effect models (p=0.018 and 0.009, respectively). Both models suggest an approximately 40% increase in the odds of treatment response [Random effects: OR=1.39 (1.06, 1.82); Fixed effect: OR=1.38 (1.08, 1.77)].

Figure 3: HAMD-17 Responder Analysis Using Random Effects Model (Responder defined as a subject with ≥ 50% reduction in baseline HAMD-17 total score at any post-baseline assessment)

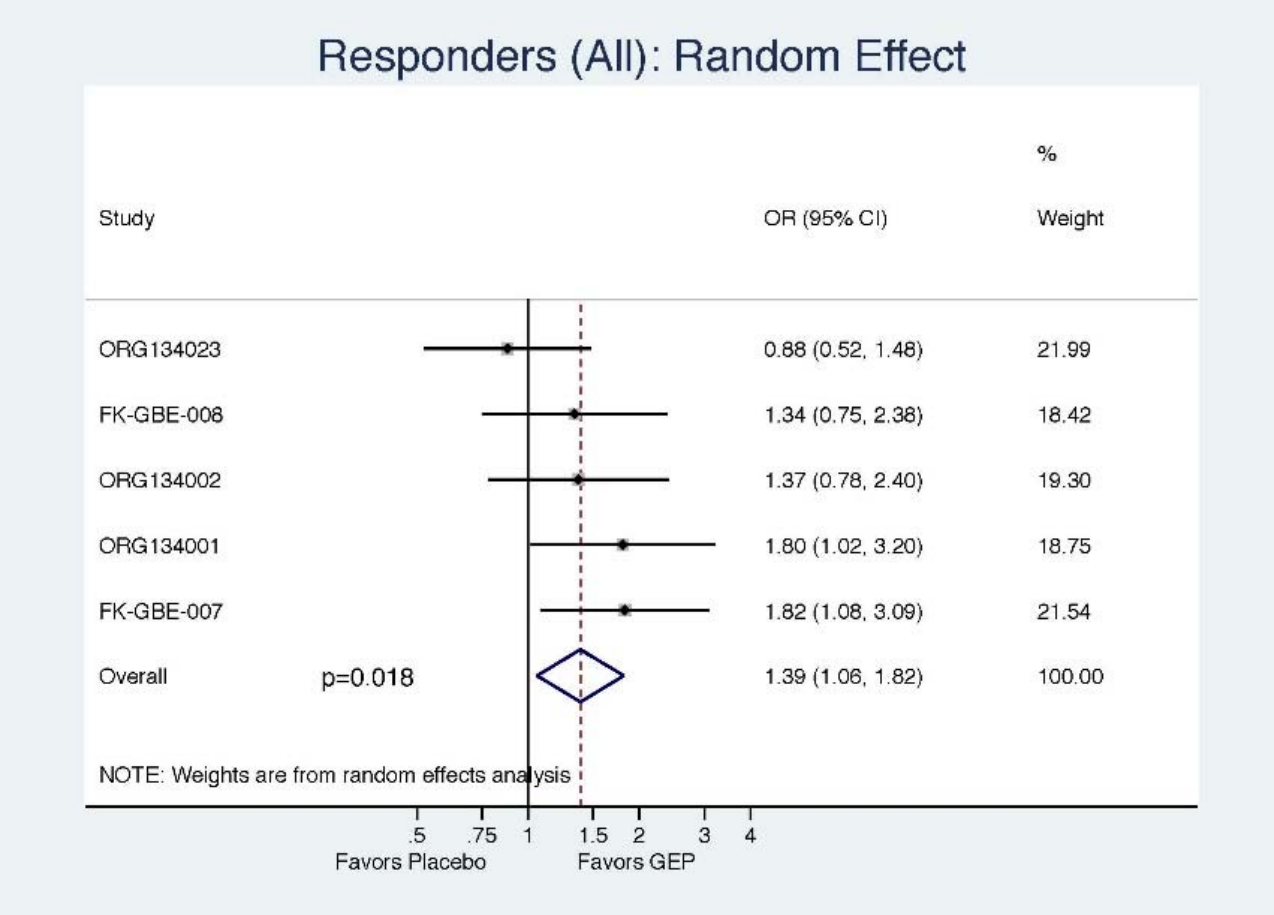


Figure 4: HAMD-17 Responder Analysis Using Fixed Effect Model (Responder defined as a subject with ≥ 50% reduction in baseline HAMD-17 total score at any post-baseline assessment)

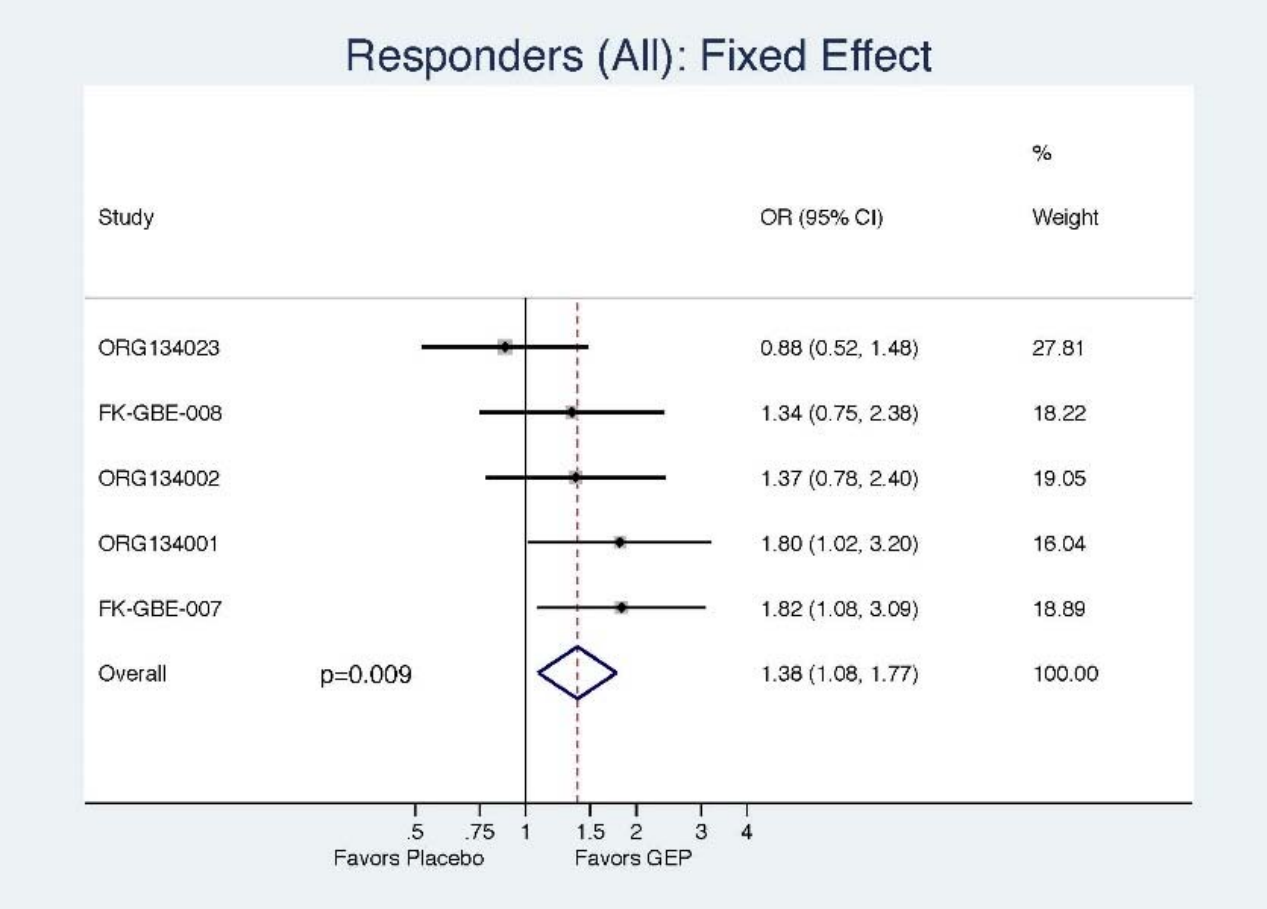


Figure 5 and Figure 6 present results of the meta-analyses for the percent of HAMD-17 responders at the last scheduled visit time point. This analysis made the pessimistic assumption that all drop-outs were non-responders, and provides an indication of the robustness of the total of the evidence supporting the effectiveness of gepirone ER in the treatment of MDD.

The meta-analysis showed a statistically significant, 42% treatment difference in favor of gepirone ER for both the random effects and fixed effect models (p=0.042 and 0.009, respectively). These results were consistent via sensitivity analyses (Figure 5 and

Figure 6) using the more pessimistic definition of treatment response at the scheduled end of the short-term, double-blind treatment period [Random effects: OR=1.42 (1.01, 1.98); Fixed effect: OR=1.41 (1.09, 1.82)].

Figure 5: HAMD-17 Responder Analysis Using Random Effects Model (Sensitivity Analysis)

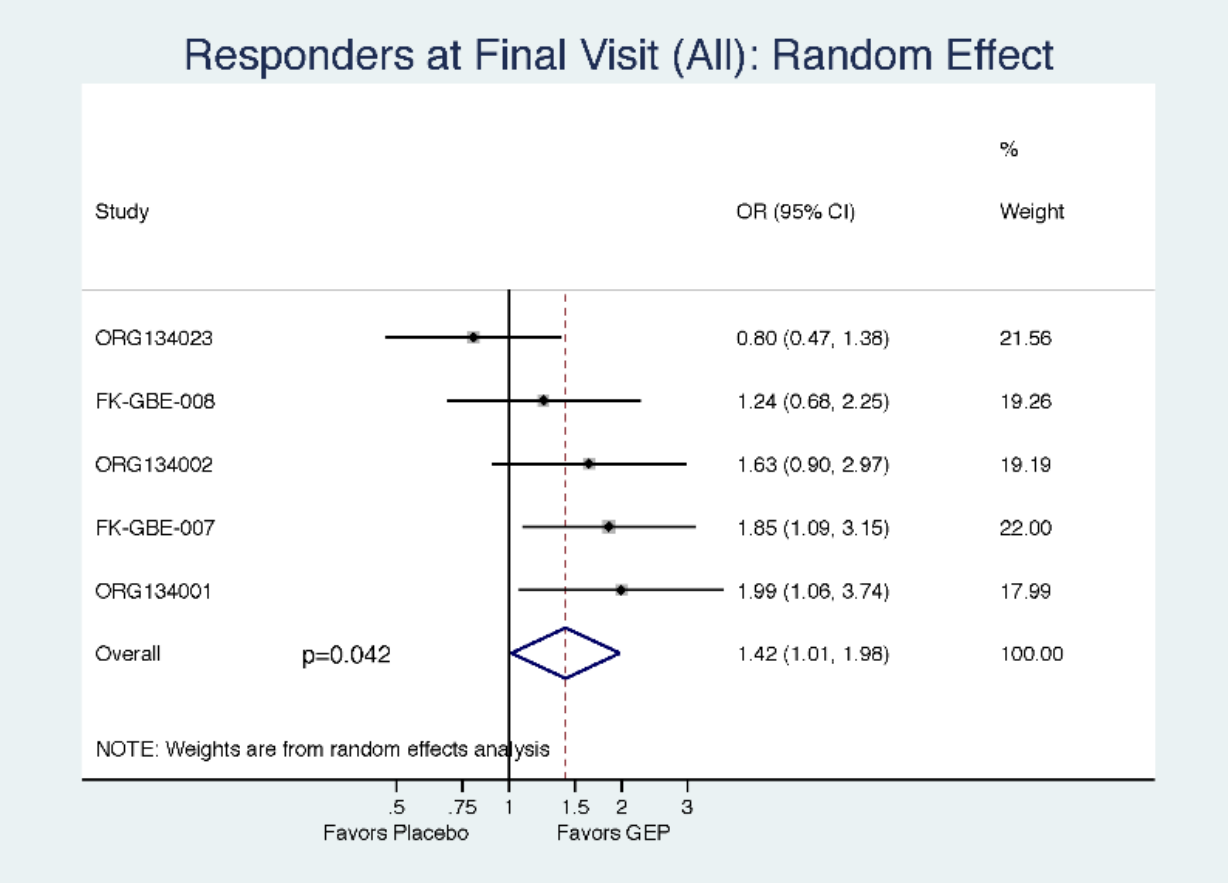
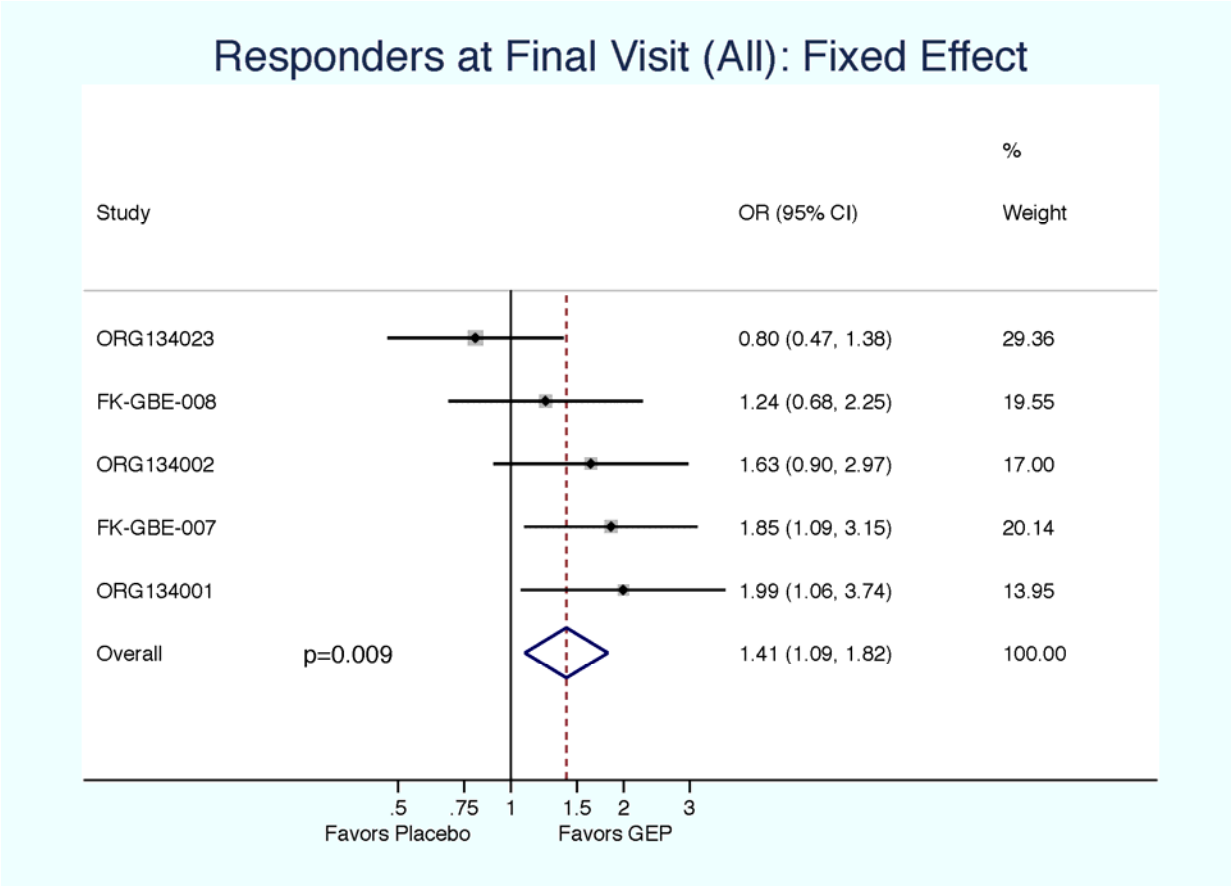


Figure 6: HAMD-17 Responder Analysis Using Fixed Effect Model (Sensitivity Analysis)



3.6.3 Conclusions from Meta-Analyses

The meta-analyses described above utilized all data from the five randomized trials that met the eligibility criteria for studies of gepirone ER in the treatment of MDD.

The meta-analyses of the primary efficacy parameter (adjusted mean change in HAMD-17 from baseline to end of treatment), were statistically significant and clinically meaningful in favor of gepirone ER compared to placebo using both random effects and fixed effect models.

The meta-analyses of the HAMD-17 responder rates, which provide a measure of clinical benefit, were also statistically significant, representing large treatment differences in favor of gepirone ER compared to placebo using both random effects and fixed effect models. Importantly, this effect was maintained when all drop-outs were considered as non-responders, which is a more pessimistic assumption than the definition used in the pre-specified analysis from the original studies.

Overall, these meta-analyses provide complementary support for the findings from the individual interpretable studies by pooling data and providing a larger sample size. Results of these meta-analyses provide further evidence that gepirone ER is effective in the treatment of MDD.

3.7 Additional Efficacy Data

One long-term Relapse Prevention study (28709) was also conducted (a detailed description of the study is provided in Appendix 3). Notably, positive efficacy data from relapse prevention studies are not a pre-approval requirement for antidepressants in the treatment of MDD, and deficiencies in 28709 are similar to those cited by FDA in its review of Fetzima (levomilnacipran), which was approved and also had a negative relapse prevention study.

Study 28709 was a multicenter, placebo-controlled trial in MDD outpatients. The primary objective of the trial was to compare the relapse rates of depression during the continuation phase between subjects receiving gepirone ER at the final titrated dose and subjects receiving placebo. The trial started with an open-label (OL) gepirone ER treatment phase of 8-12 weeks. Responders, those reaching a HAMD-17 of 8 or less, were entered into a double blind continuation phase of 40-44 weeks.

Of the 303 subjects who completed the OL phase, 250 subjects met responder criteria and were randomized into the double-blind continuation phase and treated with trial medication. The relapse rate at endpoint of the continuation phase (based on the primary analysis of the primary efficacy parameter) was 23.0% for subjects in the gepirone ER group compared to 34.7% for subjects in the placebo group (p=0.024). Beginning at Week 16, a difference in relapse rates became apparent; from Week 24 onward, the differences were statistically significant in favor of gepirone ER. When 5 subjects who discontinued the study for other reasons were included as relapses, trends were similar but differences were significant only at Week 28 (p=0.032).
As discussed in Appendix 3, numerous subjects randomized to the double-blind period were entered outside of the 8-12 week window or had other protocol violations. Eliminating these subjects in a 'per protocol' analysis, the difference in relapse rates is statistically significant favoring gepirone ER. FKP believes that only true responders are appropriate for the study (i.e., response to gepirone ER should have been confirmed on more than 1 visit prior to randomization). Unfortunately, subjects were randomized as soon as HAMD scores reached 8. Eleven subjects qualified for randomization (HAMD-17 score 8 or less) at one visit and relapsed (HAMD-17 score 16 or more) at the very next visit.

The difference between treatment groups in time to relapse did not achieve statistical significance based on the log-rank test, with p=0.065.

Overall, study 28709 detects trends that are of supportive, but deficiencies in the study might have caused larger variability than expected and a somewhat diminished effect size than expected. The failure of this study to generate positive findings on its primary endpoint should not be taken as evidence that gepirone ER is ineffective.

3.8 Overall Efficacy Summary and Conclusions

Of the twelve short-term gepirone ER clinical studies that have been conducted, five are considered well-controlled, adequately powered, randomized clinical trials that are interpretable for efficacy. Of the five interpretable studies, two clinical trials demonstrate gepirone ER's effectiveness on the primary endpoint: reduction in HAMD-17 score with effect sizes comparable to those of many approved antidepressants. In addition, two of the remaining three studies demonstrate gepirone ER's effectiveness on secondary endpoints, including: response rate as defined as the proportion of subjects with a 50% or greater reduction in HAMD-17 score. Therefore, two of the five interpretable studies provide well-demonstrated evidence of effectiveness and four of the five interpretable studies provide either well-demonstrated or supportive evidence of effectiveness. One of the five interpretable studies was considered adequately designed and provided no evidence of antidepressant efficacy.

The efficacy of gepirone ER based on all five interpretable studies was confirmed in a meta-analysis. This meta-analysis demonstrated statistically significant findings in favor of gepirone ER compared to placebo using both random effects and fixed effect models for two efficacy endpoints: (1) reduction in HAMD-17 from baseline to end of treatment; and (2) HAMD-17 responder rates.

Seven short-term studies were considered uninterpretable for efficacy and therefore were not included in our assessment for efficacy. This is consistent with standard statistical rules and FDA review policies that have been applied to efficacy reviews of other antidepressant NDAs.

4. SAFETY

4.1 Safety Profile

4.1.1 Gepirone ER - Overall Safety Assessment

Overall, the safety profile of gepirone ER is similar to that of other antidepressants, and extensive safety analyses have been conducted across gepirone ER trials that span over 20 years by three different sponsors. These studies provides a comprehensive safety database that supports the overall safety of gepirone ER.

The safety data in clinical populations is presented using all subjects treated in gepirone ER controlled Phase II/III studies for depression (n=19) (Data summarized in Table 5). Across these studies, 1,976 subjects received gepirone ER, 1,275 received placebo, 595 received fluoxetine, 276 received paroxetine, and 74 received imipramine. The percentages of subjects with at least one adverse event (AE) was highest in the imipramine group (95.9%) and ranged from 75.5% (fluoxetine) to 83.7% (gepirone ER) in the remaining treatment groups.

The body systems most frequently affected were the nervous system and gastrointestinal system (gepirone ER: 55.7% and 45.6%, respectively; placebo: 40.5% and 36.1%; fluoxetine: 34.3% and 38.2%; paroxetine: 40.2% and 43.1%; and imipramine: 83.8% and 85.1%; respectively).

AEs reported by at least 5% of all subjects included headache, dizziness, somnolence, nausea, dry mouth, diarrhea, constipation, insomnia, upper respiratory tract infection, nasopharyngitis, and fatigue. Of these, transient and dosing-related dizziness (31.0% vs. 9.8%), headache (28.3% vs. 24.3%), nausea (26.5% vs. 11.8%), insomnia (11.9% vs. 7.7%), and fatigue (5.7% vs. 4.7%) occurred in a higher proportion of subjects receiving gepirone ER than placebo.

Available information with regard to suicidality associated with gepirone ER is consistent with information contained in the class warning in US labeling for all antidepressants. The recommended FDA analyses (i.e. Columbia) have been conducted and indicate that treatment with gepirone ER does not statistically significantly increase the risk of suicidal behavior or suicidal ideation relative to placebo.

Nine deaths occurred in patients who participated in Phase II/III studies. Five of these patients took either gepirone IR or ER. None of these deaths were deemed to be drug related.

Table 5: Number (%) of Subjects with AEs Regardless of Relationship to Study Drug by Preferred Term and Treatment for AEs Reported by ≥5% of Subjects in any Treatment Group (All Subjects Treated in Gepirone ER Controlled Phase II/III Studies in Depression)

Body System			Antidepressants		
Preferred Term	Gepirone ER (N=1976)	Placebo (N=1275)	Fluoxetine (N=595)	Paroxetine (N=276)	Imipramine (N=74)
Subjects with at least 1 AE	**1653 (83.7%)**	**992 (77.8%)**	**449 (75.5%)**	**223 (80.8%)**	**71 (95.9%)**
Nervous System Disorders	**1100 (55.7%)**	**517 (40.5%)**	**204 (34.3%)**	**111 (40.2%)**	**62 (83.8%)**
Headache	559 (28.3%)	310 (24.3%)	121 (20.3%)	44 (15.9%)	34 (45.9%)
Dizziness	612 (31.0%)	125 (9.8%)	36 (6.1%)	48 (17.4%)	29 (39.2%)
Somnolence	125 (6.3%)	77 (6.0%)	24 (4.0 %)	27 (9.8%)	23 (31.1%)
Gastrointestinal Disorders	**901 (45.6%)**	**460 (36.1%)**	**227 (38.2%)**	**119 (43.1%)**	**63 (85.1%)**
Nausea	524 (26.5%)	151 (11.8%)	95 (16.0%)	53 (19.2%)	22 (29.7%)
Dry Mouth	139 (7.0%)	91 (7.1%)	38 (6.4%)	36 (13.0%)	56 (75.7%)
Diarrhea	159 (8.0%)	111 (8.7%)	64 (10.8%)	19 (6.9%)	5 (6.8%)
Constipation	96 (4.9%)	63 (4.9%)	15 (2.5%)	26 (9.4%)	30 (40.5%)
Psychiatric Disorders	**572 (28.9%)**	**266 (20.9%)**	**160 (26.9%)**	**99 (35.9%)**	**34 (45.9%)**
Insomnia	235 (11.9%)	98 (7.7%)	56 (9.4%)	23 (8.3%)	6 (8.1%)
Infections and Infestations	**450 (22.8%)**	**313 (24.5%)**	**134 (22.5%)**	**62 (22.5%)**	**13 (17.6%)**
Upper Respiratory Tract Infection	124 (6.3%)	94 (7.4%)	26 (4.4%)	15 (5.4%)	2 (2.7%)
Nasopharyngitis	110 (5.6%)	70 (5.5%)	46 (7.7%)	14 (5.1%)	4 (5.4%)
General Disorders & Administrative Site Conditions	**345 (17.5%)**	**204 (16.0%)**	**92 (15.5%)**	**46 (16.7%)**	**25 (33.8%)**
Fatigue	112 (5.7%)	60 (4.7%)	45 (7.6%)	19 (6.9%)	8 (10.8%)

Dizziness and nausea were typically mild to moderate in severity and were transient. In studies that evaluated dizziness and nausea most closely (i.e. in the Phase I setting), the median duration for these adverse reactions was 1 and 2 days, respectively. In the Phase II/III setting, the incidence of new complaints of dizziness and nausea declined towards the placebo rate in the first 4 to 6 weeks of treatment. While dizziness appeared to be more likely at higher doses, a dose-response relationship was not consistently evident for nausea. No AEs of dizziness were coded as serious adverse events (SAEs). In the placebo controlled Phase II/III studies, 2.5% of gepirone ER patients and 0.5% of placebo patients, respectively, withdrew prematurely due to dizziness.

Antidepressant-related AEs that occurred in at least 2% of subjects treated with at least 40 mg/day gepirone ER in controlled Phase II/III studies in depression are presented in Table 6.

Table 6: Incidence of Antidepressant-Related AEs in All Subjects Treated in Gepirone ER Controlled Phase II/III Studies in Depression by Relationship to Study Drug and Treatment

			Antidepressants		
Body System Preferred Term Relation	Gepirone ER (N=1976) c/n (%)	Placebo (N=1275) c/n (%)	Fluoxetine (N=595) c/n (%)	Paroxetine (N=276) c/n (%)	Imipramine (N=74) c/n (%)
Any Adverse Event					
Unrelated	17/17 (0.9%)	8/7 (0.5%)	1/1 (0.2%)	1/1 (04%)	0/0
Related	74/57 (2.9%)	62/42 (3.3%)	98/63 (10.6 %)	76/49 (17.8%)	29/18 (24.3%)
Psychiatric Disorders Any Adverse Event					
Unrelated	6/6 (0.3%)	6/5 (0.4%)	0/0	1/1(0.4%)	1/1 (1.4%)
Related	51/40 (2.0%)	48/31 (2.4%)	72/48 (8.1%)	53/36 (13.0%)	9/7 (9.5%)

Source: ISS Table 5.1.4
Adverse events (AEs) reported by 2% or more of subjects receiving at least 40 mg/day of gepirone ER.
Note: c=number of AEs reported, n=number of subjects reporting AEs, (%) = percentage of subjects reporting AEs. Subjects reporting the same AE more than once are counted only once at the greatest relationship to study drug.

The percentages of subjects with at least one antidepressant-related AE considered to be related to the study drug was comparable in the gepirone ER (2.9%) and placebo (3.3%) groups, and less than that in other antidepressant treatment groups (imipramine, 24.3%; paroxetine, 17.8%; fluoxetine, 10.6%). Results were similar within the psychiatric body system.

Long-term studies with gepirone ER have not revealed any new safety concerns. Various long-term studies have exposed more than 1,500 subjects to gepirone ER. In long-term studies, 692 subjects were treated with gepirone ER for over 6 months and 170 subjects were treated for over 1 year. In general, the extension phase and long-term (at least 6 months and at least 1 year of treatment) safety results are consistent with those reported during the acute phase.

The available human safety data for gepirone ER are extensive. The data presented here, in addition to other studies, provide a comprehensive safety database that support the overall safety of gepirone ER. Overall, gepirone ER is well tolerated among subjects with MDD, with dizziness being the most common side effect, as is typical with 5-HT1$_A$ agonists (Wilson 2005).

4.1.2 Sexual Adverse Events

Adverse events related to sexual function were identified and classified (only in Phase 2/3 studies in depression that used a sexual dysfunction questionnaire). Table 7 presents adverse event incidence for each of the categories.

Table 7: Incidence of Sexual Dysfunction Adverse Events: All Gepirone ER Phase II/III Studies in Depression

Body System/ Preferred Term		Number of Patients Experiencing AE (% of N)				
		Gepirone ER (N=1976)	Placebo (N=1275)	Active Comparators		
				Fluoxetine (N=595)	Paroxetine (N=276)	Imipramine (N=74)
Any Sex-Related Adverse Event*		47 (2.4%)	40 (3.1%)	61 (10.3%)	47 (17.0%)	10 (13.5%)
Psychiatric Sexual Disorders	Any adverse event	38 (1.9%)	34 (2.7%)	48 (8.1%)	35 (12.7%)	7 (9.5%)
	Libido decreased	21 (1.1%)	22 (1.7%)	30 (5.0%)	15 (5.4%)	4 (5.4%)

Source: ISS Table 14.1.2.1
n = number of subjects with one or more occurrence of the adverse event, expressed as a % of the group total (N)
**MedDRA High Level Group Terms: sexual function and fertility disorders, orgasmic disorders and disturbances, sexual arousal disorders, sexual desire disorders.*

Pooled safety data from all Phase II/III studies of gepirone ER in depression showed a lower incidence of sexual dysfunction with gepirone ER than placebo or SSRIs. The incidence of any sex-related adverse event was lower in subjects treated with gepirone ER (2.4%) compared to placebo (3.1%) and each of the active comparators: fluoxetine (10.3%), paroxetine (17%), and imipramine (13.5%).

5. RISK/BENEFIT PROFILE IN THE TREATMENT OF MDD

5.1.1 Additional Clinical Utility of Gepirone ER - Sexual Functioning

MDD is often accompanied by sexual dysfunction (i.e., loss of desire, diminished sexual interest, reduced sexual behavior, lessened arousal, loss of orgasm, etc.), affecting both men and women. An effective treatment of MDD would hopefully possess the capacity to restore sexual functioning towards normal. Unfortunately, many antidepressants, including the SSRIs and serotonin norepinephrine reuptake inhibitors (SNRIs) not only fail to relieve, but, on their own, cause significant sexual dysfunction (Higgins 2010) (Reviewed in more detail in Section 2.1.1).

Among the 5 interpretable studies for efficacy, two studies, 134001 and 134002, measured sexual functioning through administered standardized validated sexual functioning questionnaires. The other three interpretable studies did not measure sexual functioning.

Specifically, in studies 134001 and 134002, the Derogatis Interview for Sexual Function – Self-Report (DISF-SR) was administered to subjects (Derogatis and Melisaratos 1979). Briefly, this validated instrument is designed to provide an estimate of the quality of an individual's current sexual functioning, and gender-specific versions of the DISF-SR exist for both males and females. DISF items are arranged into 5 primary domains of sexual functioning: sexual cognition/fantasy, sexual arousal, sexual behavior/experience, orgasm and sexual drive/relationship. In addition, an aggregate DISF total score is computed which summarizes quality of sexual functioning across the five primary DISF domains.

The following sections describe the sexual functioning results from 134001 and 134002 in more detail. Note that there was no active control in either of these studies, and so corresponding comparisons were not possible.

5.1.1.1 **Sexual Functioning Results from 134001**

Female subjects receiving gepirone ER increased their mean Week 8 total DISF-SR score by 11.0 points, while their placebo counterparts saw an increase of 0.3 points. Gepirone ER-treated females experienced improvements from baseline in 29 of 30 individual mean scores at Week 8 (1 mean score remained unchanged from baseline), versus an improvement for placebo-treated females in mean scores for 12 items (8 items showed worsening from baseline at Week 8, and 10 items had scores unchanged from mean baseline scores).

Gepirone ER-treated males had an improvement in total mean item score at Week 8 of 13.7 points, compared to an increase of 0.9 points for placebo subjects. The gepirone ER group also self-reported improved Week 8 mean scores compared to baseline in 28 of 30 items (one item had a score unchanged from baseline, and one item showed a worsening from mean baseline score at Week 8). The placebo-treated males reported improvements in 17 individual item scores, worsening of mean scores in 12 items, and 1 score unchanged from baseline at Week 8.

An analysis of variance (ANOVA) was performed for males and females combined for the change from baseline to the end of study in DISF total score. The difference between placebo

and gepirone ER (placebo minus gepirone ER) for change from baseline in DISF-SR total score was (-11.2; CI: -19.5, -3.0).

These data indicate that gepirone ER subjects had statistically better sexual functioning than placebo subjects as measured by the DISF-SR.

5.1.1.2 **Sexual Functioning Results from 134002**

Female subjects in both treatment groups experienced an increase in their total Week 8 mean item scores over baseline (gepirone ER: 3.6 points; placebo: 9.9 points). Females treated with gepirone ER saw improvements from baseline in 18 of 30 individual scores (7 items had worsening of Week 8 mean scores compared to baseline, and scores for 5 items remained unchanged from baseline). Females taking placebo improved their Week 8 mean scores from baseline in 28 items, while 2 scores remained unchanged from baseline.

Gepirone ER-treated males saw an increase in their total mean Week 8 score of 13.6 points, compared to a decrease (worsening) in total mean (SD) score within the placebo group of 3.1 points (-3.1). For the individual items, those males receiving gepirone ER experienced an improvement in mean scores at Week 8 in 19 items, and had a worsening from baseline mean scores for 6 items. Mean scores for 5 of the 30 items remained unchanged from baseline at Week 8. In the placebo group, there were self-reported improvements at Week 8 for mean scores in 6 of 30 items. The male placebo subjects also experienced worsening from mean baseline scores in 20 items, and saw mean scores at Week 8 remain unchanged from baseline for 4 of the items.

An ANOVA was performed for males and females combined for the change from baseline to the end of study in DISF total score. Gepirone ER subjects showed a greater improvement from baseline in DISF-SR total score compared with placebo (-1.9; CI: -10.5, 6.7).

5.1.2 Conclusions on Sexual Functioning with Gepirone ER

Sexual functioning was assessed in two of the five interpretable studies for efficacy using the DISF-SR validated scale.

The results indicate that gepirone ER-treated subjects had much better sexual functioning scores at the end of study 134001 compared to placebo-treated subjects on the DISF-SR. In study 134002, gepirone ER subjects showed more improvement from baseline in DISF-SR total score compared with placebo.

As mentioned earlier, studies 134001 and 134002 did not have active comparator antidepressant groups. However, as discussed in section 4.1.2, an evaluation of sexual functioning related adverse events indicated that the risk of developing any sexual dysfunction-related AE in subjects treated with gepirone ER was similar to placebo, while comparator drugs, fluoxetine, paroxetine, and imipramine were all associated with an increased risk of sexual dysfunction AEs, relative to placebo or gepirone ER.

Overall, sexual dysfunction is highly prevalent in MDD, and is estimated to be greater than 70% in this population. An effective treatment of MDD would possess the capacity to restore sexual functioning towards normal. Many currently available antidepressants, including the SSRIs and SNRIs not only fail to relieve, but, on their own, cause increased sexual dysfunction. Based on gepirone ER's mechanism of action, and demonstrated by its clinical and safety data, gepirone ER may provide additional clinical utility with regard to sexual functioning both by not impairing sexual function in depressed patients who do not have sexual dysfunction pre-treatment, and by improving sexual function in depressed patients who do have sexual dysfunction pre-treatment.

5.2 Further Support for Sexual Functioning Benefits

Data from other gepirone ER studies were consistent with the results seen in 134001 and 134002. In the overall gepirone ER development program, sexual function was assessed in 5 short-term studies, one 52 week relapse prevention study, and three long term extension studies. While three of the short term studies were not interpretable for efficacy parameters, safety results were utilized from all studies. These overall results were consistent across studies and various rating scales and diagnostic criteria.

Analysis of DISF and Changes in Sexual Functioning Questionnaire (CSFQ) scores as well as DSM- IV assessments from pooled studies revealed the following:

- Sexual dysfunction is highly prevalent in depression (73% in this study population).
- The severity of sexual dysfunction in depressed subjects is highly correlated with the severity of depression.
- Gepirone-ER significantly improved sexual function in subjects with pre-treatment sexual dysfunction relative to placebo and maintained normal levels of sexual function in those without pre-treatment sexual dysfunction. Greater improvement was observed with more severe pre-treatment sexual dysfunction.
- Fluoxetine and paroxetine had a negative effect on sexual function; subjects who started these drugs with normal sexual function experienced a significant decline in sexual function, whereas those with pre-treatment sexual dysfunction showed no improvement.
- Gepirone-ER improved sexual function in both males and females, most notably the DISF domains of desire, drive, and orgasm.

5.3 Risk/Benefit Profile

The benefits of gepirone ER are measured primarily in terms of efficacy as evidenced in five interpretable clinical trials. In addition, gepirone ER has sexual functioning benefits, which can improve the quality of life for MDD patients, given that sexual dysfunction is highly prevalent in depression and estimated to affect more than 70% of this patient population.

Overall, gepirone ER provides the following specific benefits to MDD patients:

- Effective in the treatment of MDD
- Clean safety profile, comparable, or better, relative to other antidepressants
- No treatment-induced sexual dysfunction

The risks of gepirone ER are measured primarily in terms of its safety profile, in addition to non-clinical data. As discussed earlier, the available human safety data for gepirone ER is extensive, and provides a comprehensive safety database that supports the overall safety of gepirone ER.

The common risks occurring at an incidence >5% and that are related to gepirone ER treatment are AEs that may be unpleasant to patients but have not resulted in medically severe sequelae. The AEs which occurred in a higher proportion of subjects receiving gepirone ER than placebo include headache, dizziness, nausea, insomnia, and fatigue.

With regard to key findings from a standard battery of non-clinical studies, gepirone ER was not found to have mutagenic, teratogenic, or carcinogenic potential. No adverse effects of gepirone ER on fertility were observed. Gepirone ER is capable of eliciting signs of serotonin syndrome in rats when administered at high doses, and preclinical data indicate that animals will not self-administer gepirone ER. In addition, there is no clinical evidence to suggest abuse in humans, no increased risk of suicide (relative to placebo), and no cardiovascular AEs.

6. DISCUSSION AND CONCLUSIONS

Gepirone ER is a new chemical entity being developed by FKP for short-term treatment of MDD. This briefing book reviews the evidence supporting the effectiveness of gepirone ER and is intended to provide the Committee with the data and analyses needed to address the questions posed by the Agency with respect to the approvability of this product for the treatment of MDD.

The novel mechanism of action of gepirone ER (agonist activity at the 5-HT1$_A$ receptor) sets it apart from currently approved antidepressant medications and makes it a unique chemical entity in the treatment of MDD. While other approved antidepressants are partial 5-HT1$_A$ agonists, these agents also serve as serotonin uptake inhibiters and therefore produce widespread increases in extracellular serotonin. Because the mechanism of action gepirone ER appears to be limited to 5-HT1$_A$ agonism (without serotonin uptake inhibition), its side effect profile is devoid of significant sexual adverse effects.

FKP considers five of the twelve well-controlled, adequately powered, randomized clinical trials to be interpretable and therefore should serve as the basis for an assessment of the efficacy of gepirone ER. Of the five interpretable studies, two clinical trials provide strong and statistically significant evidence of gepirone ER's effectiveness on the primary endpoint: reduction in HAMD-17 score, and exhibit an effect size comparable to many other approved antidepressants. Based on statistical significance on some secondary endpoints, two other interpretable studies are also supportive of efficacy.

Of the twelve short-term gepirone ER clinical studies that have been conducted, FKP considers seven to be uninterpretable due to a series of statistical and methodological principles. Because these 7 trials are uninterpretable, they should not be considered part of the dataset to evaluate the effectiveness of gepirone ER.

A series of meta-analyses were designed and performed on all five interpretable studies as a means to assess the totality of evidence. These analyses demonstrated statistically significant benefits of gepirone ER on the following endpoints taking into account all five interpretable studies: (1) reduction in HAM-D score; and (2) response rate as defined as greater than 50% reduction in HAMD-17 score. Taken together with the findings from the two individual studies demonstrating efficacy for gepirone ER and the additional supportive data, this meta-analysis provides compelling evidence of the benefits of gepirone ER in MDD.

Seven short-term studies were considered uninterpretable and therefore were not included in the assessment for efficacy, and one long-term study was excluded because of its different design.

While safety is not a core issue to be addressed at this Advisory Committee meeting, it should be noted that gepirone ER has a favorable safety profile. Based on this favorable safety profile, gepirone ER's benefits appear to outweigh its risks. Specifically, based largely on the observation of a beneficial safety profile with respect to sexual dysfunction side effects, gepirone ER may offer clinical advantages over other currently approved serotonergic antidepressants.

Taken as a whole, data from five well-controlled, adequately powered, randomized, interpretable studies provide substantial evidence of effectiveness for gepirone ER in the treatment of major depressive disorder. The data from this program also provide evidence that gepirone ER is well tolerated and safe for patients with MDD.

7. REFERENCES

Amsterdam, J.D. 1992. Gepirone, a selective serotonin (5-HT1$_A$) partial agonist in the treatment of major depression. *Prog.Neuropsychopharmacol.Biol.Psychiatry.* 16:271-280.

Angst, J., and M. Preisig. 1995. Course of a clinical cohort of unipolar, bipolar and schizoaffective patients. Results of a prospective study from 1959 to 1985. *Schweiz.Arch.Neurol.Psychiatr.* 146:5-16.

Angst, J., and M. Preisig. 1995. Outcome of a clinical cohort of unipolar, bipolar and schizoaffective patients. Results of a prospective study from 1959 to 1985. *Schweiz.Arch.Neurol.Psychiatr.* 146:17-23.

Bishop JR, Moline J, Ellingrod VL, Schultz SK, Clayton AH. 2006. Serotonin 2A -1438 G/A and G-protein Beta3 subunit C825T polymorphisms in patients with depression and SSRI-associated sexual side-effects. Neuropsychopharmacology. Oct;31(10):2281-8.

Centers for Disease Control and Prevention. 2015. 10 Leading Causes of Death by Age Group, United States – 2013. 2015.

Clayton A., A. Keller, and E.L. McGarvey. 2006. Burden of phase-specific sexual dysfunction with SSRIs. Journal of Affective Disorders. 91:27-32.

Csoka, A.B., A. Bahrick, and O.P. Mehtonen. 2008. Persistent sexual dysfunction after discontinuation of selective serotonin reuptake inhibitors. *J.Sex.Med.* 5:227-233.

Derogatis L.R., and N. Melisaratos. 1979. The DSFI: a multidimensional measure of sexual functioning. J Sex Marital Ther. Fall;5(3):244-81.

DerSimonian, R., and N. Laird. 1986. Meta-analysis in clinical trials. Controlled clinical trials, 7(3), 177-188.

Dunlop, B.W., and M.H. Rapaport. 2015. Antidepressant signal detection in the clinical trials vortex. *J.Clin.Psychiatry.* 76:e657-9.

Evans, D.L., D.S. Charney, L. Lewis, R.N. Golden, J.M. Gorman, K.R. Krishnan, C.B. Nemeroff, J.D. Bremner, R.M. Carney, J.C. Coyne, M.R. Delong, N. Frasure-Smith, A.H. Glassman, P.W. Gold, I. Grant, L. Gwyther, G. Ironson, R.L. Johnson, A.M. Kanner, W.J. Katon, P.G. Kaufmann, F.J. Keefe, T. Ketter, T.P. Laughren, J. Leserman, C.G. Lyketsos, W.M. McDonald, B.S. McEwen, A.H. Miller, D. Musselman, C. O'Connor, J.M. Petitto, B.G. Pollock, R.G. Robinson, S.P. Roose, J. Rowland, Y. Sheline, D.S. Sheps, G. Simon, D. Spiegel, A. Stunkard, T. Sunderland, P. Tibbits Jr, and W.J. Valvo. 2005. Mood disorders in the medically ill: scientific review and recommendations. *Biol.Psychiatry.* 58:175-189.

Fabre, L.F., A.H. Clayton, L.C. Smith, I. Goldstein, and L.R. Derogatis. 2012. The effect of gepirone ER in the treatment of sexual dysfunction in depressed men. *J.Sex.Med.* 9:821-829.

Fabre, L.F., and L.C. Smith. 2012. The effect of major depression on sexual function in women. *J.Sex.Med.* 9:231-239.

Fabre, L.F., L.C. Smith, and L.R. DeRogatis. 2011. Gepirone ER treatment of low sexual desire associated with depression in women as measured by the DeRogatis Inventory of Sexual Function (DISF) fantasy/cognition (desire) domain--a post hoc analysis. *J.Sex.Med.* 8:2569-2581.

FDA. 1998. Guidance for Industry – Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products. U.S. Department of Health and Human Services. FDA. Clinical 6.

Fitton, A., and P. Benfield. 1994. Gepirone in depression and anxiety disorders. *CNS Drugs.* 1:388-398.

Furukawa, T.A., R.C. Kessler, T. Slade, and G. Andrews. 2003. The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. *Psychol.Med.* 33:357-362.

Greenberg, P.E., A.A. Fournier, T. Sisitsky, C.T. Pike, and R.C. Kessler. 2015. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *J.Clin.Psychiatry.* 76:155-162.

Harsh V, McGarvey EL, Clayton, AH. Physician Attitudes Regarding Hypoactive Sexual Desire Disorder in a Primary Care Clinic: A Pilot Study. J Sex Med 2008; 5:640-645

Higgins A, Nash M, Lynch AM. 2010. Antidepressant-associated sexual dysfunction: impact, effects, and treatment. Drug Healthc Patient Saf; 2:141-50.

Isacsson, G. 2006. Depression is the core of suicidality--its treatment is the cure. *Acta Psychiatr.Scand.* 114:149-150.

Israel, J. 2010. The Impact of Residual Symptoms in Major Depression. Pharmaceuticals 2010, 3(8), 2426-2440

Kennedy, S.H., and S. Rizvi. 2009. Sexual dysfunction, depression, and the impact of antidepressants. *J.Clin.Psychopharmacol.* 29:157-164.

Kessler, R.C., P. Berglund, O. Demler, R. Jin, D. Koretz, K.R. Merikangas, A.J. Rush, E.E. Walters, P.S. Wang, and National Comorbidity Survey Replication. 2003. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *Jama.* 289:3095-3105.

Kessler, R.C., P. Berglund, O. Demler, R. Jin, K.R. Merikangas, and E.E. Walters. 2005. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch.Gen.Psychiatry.* 62:593-602.

Khin, N.A., Y.F. Chen, Y. Yang, P. Yang, and T.P. Laughren. 2011. Exploratory analyses of efficacy data from major depressive disorder trials submitted to the US Food and Drug Administration in support of new drug applications. *J.Clin.Psychiatry.* 72:464-472.

Kingsberg SA. 2009. Identifying HSDD in the family medicine setting. *J Fam Pract*.58(7 Suppl. Hypoactive):S22-5.

Laughren, T.P., J. Gobburu, R.J. Temple, E.F. Unger, A. Bhattaram, P.V. Dinh, L. Fossom, H.M. Hung, V. Klimek, J.E. Lee, R.L. Levin, C.Y. Lindberg, M. Mathis, B.N. Rosloff, S.J. Wang, Y. Wang, P. Yang, B. Yu, H. Zhang, L. Zhang, and I. Zineh. 2011. Vilazodone: clinical basis for the US Food and Drug Administration's approval of a new antidepressant. *J.Clin.Psychiatry.* 72:1166-1173.

Laughren, T.P. 2010. What's next after 50 years of psychiatric drug development: an FDA perspective. *J.Clin.Psychiatry.* 71:1196-1204.

Mann, J.J. 2005. The medical management of depression. *N.Engl.J.Med.* 353:1819-1834.

McGrath, P.J., J.W. Stewart, F.M. Quitkin, Y. Chen, J.E. Alpert, A.A. Nierenberg, M. Fava, J. Cheng, and E. Petkova. 2006. Predictors of relapse in a prospective study of fluoxetine treatment of major depression. *Am.J.Psychiatry.* 163:1542-1548.

Medic, G., K. Higashi, K.J. Littlewood, T. Diez, O. Granstrom, and R.S. Kahn. 2013. Dosing frequency and adherence in chronic psychiatric disease: systematic review and meta-analysis. *Neuropsychiatr.Dis.Treat.* 9:119-131.

Michaud, C.M., M.T. McKenna, S. Begg, N. Tomijima, M. Majmudar, M.T. Bulzacchelli, S. Ebrahim, M. Ezzati, J.A. Salomon, J.G. Kreiser, M. Hogan, and C.J. Murray. 2006. The burden of disease and injury in the United States 1996. *Popul.Health.Metr.* 4:11.

Nurnberg, H.G. 2008. An evidence-based review updating the various treatment and management approaches to serotonin reuptake inhibitor-associated sexual dysfunction. *Drugs Today (Barc).* 44:147-168.

Nurnberg, H.G., and P.L. Hensley. 2003. Selective phosphodiesterase type-5 inhibitor treatment of serotonergic reuptake inhibitor antidepressant-associated sexual dysfunction: a review of diagnosis, treatment, and relevance. *CNS Spectr.* 8:194-202.

Pfaus JG. 2009. Pathways of sexual desire. J Sex Med. Jun;6(6):1506-33.

Rush, A.J., M.H. Trivedi, S.R. Wisniewski, A.A. Nierenberg, J.W. Stewart, D. Warden, G. Niederehe, M.E. Thase, P.W. Lavori, B.D. Lebowitz, P.J. McGrath, J.F. Rosenbaum, H.A. Sackeim, D.J. Kupfer, J. Luther, and M. Fava. 2006. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am.J.Psychiatry.* 163:1905-1917.

Rush, A.J., M. Fava, S.R. Wisniewski, P.W. Lavori, M.H. Trivedi, H.A. Sackeim, M.E. Thase, A.A. Nierenberg, F.M. Quitkin, T.M. Kashner, D.J. Kupfer, J.F. Rosenbaum, J. Alpert, J.W. Stewart, P.J. McGrath, M.M. Biggs, K. Shores-Wilson, B.D. Lebowitz, L. Ritz, G. Niederehe, and STAR*D Investigators Group. 2004. Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Control.Clin.Trials.* 25:119-142.

Solomon, D.A., M.B. Keller, A.C. Leon, T.I. Mueller, M.T. Shea, M. Warshaw, J.D. Maser, W. Coryell, and J. Endicott. 1997. Recovery from major depression. A 10-year prospective follow-up across multiple episodes. *Arch.Gen.Psychiatry.* 54:1001-1006.

Spijker, J., R. de Graaf, R.V. Bijl, A.T. Beekman, J. Ormel, and W.A. Nolen. 2002. Duration of major depressive episodes in the general population: results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Br.J.Psychiatry.* 181:208-213.

Stahl, S.M. 1993. Serotonergic mechanisms and the new antidepressants. *Psychol.Med.* 23:281-285.

Sulzer, D. 2011. How addictive drugs disrupt presynaptic dopamine neurotransmission. *Neuron.* 69:628-649.

Timmer, C.J., and J.M. Sitsen. 2003. Pharmacokinetic evaluation of gepirone immediate-release capsules and gepirone extended-release tablets in healthy volunteers. *J.Pharm.Sci.* 92:1773-1778.

Wang, S.M., C. Han, S.J. Lee, A.A. Patkar, P.S. Masand, and C.U. Pae. 2015. Vilazodone for the treatment of major depressive disorder: focusing on its clinical studies and mechanism of action. *Psychiatry.Investig.* 12:155-163.

Weissman, M.M., R.C. Bland, G.J. Canino, C. Faravelli, S. Greenwald, H.G. Hwu, P.R. Joyce, E.G. Karam, C.K. Lee, J. Lellouch, J.P. Lepine, S.C. Newman, M. Rubio-Stipec, J.E. Wells, P.J. Wickramaratne, H. Wittchen, and E.K. Yeh. 1996. Cross-national epidemiology of major depression and bipolar disorder. *Jama.* 276:293-299.

World Health Organization. 2008. *The Global Burden of Disease: 2004 Update*. WHO Press, Geneva, Switzerland.

Wilson, S.J., J.E. Bailey, and D.J. Nutt. 2005. Dizziness produced by a potent 5HT1$_A$ receptor agonist (eptapirone) is not due to postural hypotension. *Psychopharmacology (Berl).* 179:895-896.

8. APPENDIX 1 – SUMMARIES OF 5 INTERPRETABLE STUDIES

APPENDIX 1

SUMMARIES OF 5 INTERPRETABLE STUDIES

8.1.1 134001

Study Description

This was a 5-center, randomized, double-blind, placebo-controlled study evaluating gepirone ER 20-80 mg/day over an 8-week treatment period in subjects with MDD. The study had one forced-titration step from 20-40 mg/day at Day 4, and a flexible-dose design thereafter. The mean dose (± SD) of gepirone ER was 61.05 (±12.02) mg/day.

Study Design

Randomized, double-blind, placebo-controlled, clinical study evaluating gepirone ER in the treatment of MDD. The study was conducted across 5-centers which enrolled patients diagnosed with moderate to severe depression. The study used a flexible dosage strategy and was 8 weeks in duration.

Each patient participated in a 7-day screening period designed to ensure that all eligibility criteria were met. Patients who met all eligibility requirements were randomized to receive either placebo or gepirone ER (20 mg to 80 mg) once per day in the morning for 8 weeks (56 days). Subjects were evaluated at baseline (Day 0) and Weeks 1, 2, 3, 4, 6 and 8. As indicated in Table 8, patients in the gepirone ER group initiated treatment by taking a 20 mg tablet and one placebo tablet once daily in the morning with food, followed by an increase to 40 mg daily between Days 4 and 7. The dose could be increased to 60 mg daily after Day 7, and to 80 mg daily after 14 days, in order to optimize the therapeutic response and tolerance after Visit 2 (Day 14).

Table 8: Flexible Dosing Schedule in Study 134001

Treatment Day	Dose (mg/day)
1-3	**20**
4-6	**40**
7-14	**40 or 60**
>14	**80**

Entrance Criteria

Inclusion Criteria

Key inclusion criteria were:

- Patients between 18 and 70 years of age;
- Patients with moderate to severe MDD (Diagnostic and Statistical Manual, Fourth Edition [DSM-IV]);
- Significant daily dysphoria for the past 4 weeks prior to screening;

- Hamilton Depression Rating Scale (HAMD)-17 total score of ≥ 20 at both screening and baseline assessments.

Exclusion Criteria

Key exclusion criteria were:

- At least 20% decrease on the HAMD-17 total score between screening and baseline;
- Patients who were ≥ 65 years, diagnosed with, or taking any medication for any chronic illness (other than psychiatric disorders);
- DSM-IV Axis I diagnosis other than MDD;
- DSM-IV Axis II disorder;
- History of treatment-refractory major depressive episodes defined as incomplete or no therapeutic response to 2 prior courses of at least 1 month duration;

Efficacy Assessments

Primary Efficacy Measure

The primary efficacy measure was the change from baseline on the HAMD-17 total score at study endpoint (Week 8 or last visit) for the Intention-to-Treat (ITT) patient population.

Secondary Efficacy Measures

Key secondary efficacy measures, were as follows:

- Change from baseline on:
  - HAMD-21, HAMD-25, and HAMD-28 total scores;
  - HAMD item [1] Depressed Mood;
  - Clinical Global Impression (CGI) improvement score;
  - HAMD anxiety/somatization factor (Factor 1);
  - HAMD item 12 (psychic anxiety)
  - Montgomery-Åsberg Depression Rating Scale (MADRS) total score;
  - Clinical Global Impression of Severity (CGI-S);
- Responder and Remitter analyses, conducted as follows:
  - Proportion of HAMD-17 responders (≥ 50% change from baseline on the HAMD-17 total score);
  - Proportion of HAMD-25 responders (≥ 50% change from baseline on the HAMD-25 total score,) for patients who satisfied DSM-IV criteria for MDD with atypical depression;
  - Proportion of MADRS responders (≥ 50% change from baseline on the MADRS total score);

- Proportion of Clinical Global Impression (CGI) responders ("much" or "very much" improved, according to the CGI Global Impression-Improvement [CGI-I] assessment) ;
- Proportion of remitters (HAMD-17 total score ≤ 7 on a post-baseline assessment).

Statistical Methods

For all continuous variables, analysis of variance (ANOVA) was carried out with treatment and center as factors. For categorical variables, analyses were performed using the Cochran-Mantel-Haenszel (CMH) test, adjusting for center. The primary analysis was based on the Intent-to-Treat (ITT) population using a Last Observation Carried Forward (LOCF) procedure. The ITT population included all randomized subjects who took study drug and had at least one post-baseline efficacy assessment. With LOCF, the last assessment for subjects who dropped out or missed visits was imputed for subsequent visits. Efficacy analyses were also performed using an Observed Case (OC) procedure with no imputation for missing data. This summary focuses on ITT/LOCF results.

Results

Overall, 208 subjects were randomized and treated (102 gepirone ER, 106 placebo).

As shown in Table 9, the drop-out rate was slightly higher in the gepirone ER group (27.5% vs. 23.6%), mostly due to adverse events (9.8% vs. 2.8%); other reasons, including lost to follow-up or withdrawn consent, were more frequent in the placebo group (13.7% vs. 17.0%). Four subjects in each group dropped out for lack of efficacy. A total of 204 subjects with post-baseline data comprised the ITT population (101 gepirone ER, 103 placebo).

Table 9: Subjects Discontinued by Reason in Study 134001

Number of Subjects		Treatment Group		
		gepirone-ER	Placebo	Total
Randomized		103	106	209
Treated†		102	106	208
Discontinued	Total	28 (27.5%)	25 (23.6%)	53 (25.5%)
	Adverse events	10 (9.8%)	3 (2.8%)	13 (6.3%)
	Lack of efficacy	4 (3.9%)	4 (3.8%)	8 (3.8%)
	Other reason	14 (13.7%)	18 (17.0%)	32 (15.4%)
Completed Treatment		74 (72.5%)	81 (76.4%)	155 (74.5%)

One randomized subject returned all study medication and was excluded from ITT.
[Source: Appendix F8.1.1-1 and F8.1.2-1]

Table 10 presents the change from baseline in HAMD-17 total score by visit and treatment group for the ITT population (LOCF).

Statistically significant differences between the gepirone ER and placebo treatment groups were noted for the change from baseline in the HAMD-17 total

score at Week 3 (p=0.013) and at Week 8/Endpoint (p=0.018), with an effect size of 2.29 points in favor of gepirone ER at the final visit. Marginally significant differences in favor of gepirone ER were observed for the change from baseline in HAMD-17 total score at Weeks 1, 2, 4 and 6 ($p < 0.10$).
Results based on the ITT/Observed Cases (OC) analysis were consistent with these findings.

Table 10: HAMD-17 Total Score: Change from Baseline at Each Visit (ITT/LOCF) in Study 134001

Treatment Group		Baseline	LS Mean Change from Baseline					
			Week 1	Week 2	Week 3	Week 4	Week 6	Week 8/ET
gepirone-ER	n	101	98	100	101	101	101	101
	Mean	22.73	-3.30	-5.74	-7.86	-8.23	-8.44	-9.04
	SE	0.24	0.47	0.58	0.65	0.67	0.75	0.78
Placebo	n	103	99	101	101	101	101	101
	Mean	22.75	-2.17	-4.40	-5.86	-6.78	-6.63	-6.75
	SE	0.25	0.46	0.57	0.64	0.66	0.74	0.77
Difference: (gep-ER – placebo)			-1.13	-1.34	-2.00	-1.45	-1.81	-2.29
p-value			0.052	0.059	0.013	0.078	0.051	0.018

Analysis is based on ANOVA with effects for treatment and center; ET=End of Treatment; LS=Least Squares; SE=Standard Error of the Mean
[*Source: CSR 134001 Table 12, Appendices F8.6.1.1-2 and F8.6.1.1-4*]

Table 11 presents the results for HAMD-17 at final visit (ITT/LOCF) for each center and overall based on an ANCOVA model, with factors for treatment, center, and baseline value as a covariate. This approach was used as a secondary analysis to further explore the treatment effect.

Table 11: HAMD-17 Total Score: Change from Baseline at End of Treatment, by Center (ITT/LOCF) in Study 134001

Center	Number of Subjects		Adjusted Mean Change		Treatment Diff (95% CI) gepirone-ER − Placebo	SE	p-value
	gepirone-ER	Placebo	gepirone-ER	Placebo			
Center 1	33	32	-10.46	-6.59	-3.88 (-6.82, -0.94)	1.47	0.011
Center 2	24	25	-7.84	-8.71	0.87 (-2.90, 4.63)	1.87	0.645
Center 3	25	24	-11.58	-8.31	-3.27 (-7.68, 1.15)	2.19	0.143
Center 4	15	15	-9.57	-5.03	-4.54 (-11.44, 2.37)	3.36	0.189
Center 5	4	5	-6.03	-3.77	-2.26 (-10.73, 6.22)	3.46	0.539
All Centers	101	101	-9.04	-6.57	-2.47 (-4.41, -0.53)	0.98	0.013
Treatment by Center Interaction p-value = 0.385							

Analysis is based on ANCOVA with effects for treatment, center, and baseline value as a covariate.
[*Source: Statistical Table 3.1 in the ISE*]

Center-specific results for HAMD-17 showed trends favoring gepirone ER in 4 of the 5 centers; the treatment effect in center 1 achieved statistical significance on its own (p=0.011). Across centers, the average reduction in HAMD-17 was

significantly greater in the gepirone ER group (-9.04 vs -6.57; p=0.013). The treatment by center interaction term was not statistically significant (p=0.385).

This covariate-adjusted analysis confirmed the presence of a statistically significant treatment effect favoring gepirone ER for the primary efficacy variable.

An overview of results for secondary efficacy variables at Endpoint (Week 8/End of Treatment) is presented in Table 12.

Treatment effects consistently favored gepirone ER over placebo for each of the secondary efficacy variables at all study visits. At Week 8/End of Treatment, statistically significant differences were noted for the change from baseline in HAMD-25 (p=0.007), MADRS (p=0.023), HAMD-Item 1 (depressed mood) (p=0.005), HAMD-Item 12 (psychic anxiety) (p=0.001), and HAMD-21 (p=0.021).

Responder rates based on HAMD-17 also favored gepirone ER; differences were statistically significant at Week 3 (p=0.017) and Week 4 (p=0.035), and marginally significant at Endpoint (p=0.059). The number (%) of CGI responders was consistently higher in the gepirone ER group compared to placebo at each visit, although the differences did not achieve statistical significance. Statistically significant differences between treatment groups were detected at Week 8/End of Treatment in the number (%) of HAMD-25 responders (p=0.014; 16.8% difference in favor of gepirone ER) and HAMD-17 remitters (p=0.017; 13.8% difference in favor of gepirone ER) based on the LOCF analysis for the ITT group. Results based on the LOCF and OC analyses for the ITT Group were consistent.

Table 12: Summary of Secondary Efficacy Results at Endpoint (ITT/LOCF) in Study 134001

Parameter	End of Treatment Outcome		Difference	p-value
	gepirone-ER	Placebo		
HAMD-17 Responders (%)	43.6%	30.7%	12.9%	0.059
HAMD-17 Remitters (%)	28.7%	14.9%	13.8%	0.017
HAMD-21 CFB	-10.01	-7.49	-2.51	0.021
HAMD-28 CFB	-13.27	-9.60	-3.68	0.013
MADRS CFB	-12.28	-9.22	-3.06	0.023
HAMD-Item 1 CFB	-1.16	-0.78	0.39	0.005
CGI Responders (%)	43.6%	35.6%	7.92	0.251
CGI severity CFB	-1.19	-0.79	-0.39	0.016
CGI improvement	2.82	3.10	0.29	0.072
HAMD Factor 1 CFB	-2.67	-2.08	0.58	0.124
HAMD-Item 12 CFB	-0.92	-0.49	0.43	0.001
HAMD-25 CFB	-11.57	-8.19	-3.38	0.007
HAMD-25 Responders (%)	45.5%	28.7%	16.8%	0.014

[*Source: 134001 CSR Table 21, Appendix F8.6.2*]
CFB=Change from baseline; LS means from ANOVA model, including effects for treatment and center; % responders and % remitters were compared using CMH test. For HAMD and MADRS scales, responders are subjects with ≥ 50% reduction from baseline; for CGI, responders are much or very much improved on the CGI improvement score.

Conclusions

This study was a 5-center, randomized, double-blind, placebo-controlled study, which employed doses in the appropriate therapeutic range for gepirone ER. Study 134001 was adequately designed and executed, and the primary efficacy parameter was the change from baseline in HAMD-17 total score based on the LOCF analysis of the ITT population.

The treatment effect was statistically significant favoring gepirone ER over placebo for the primary efficacy variable, and this positive result was supported by nearly all secondary efficacy variables.

Study 134001 provides strong evidence of the therapeutic effectiveness of gepirone ER for the treatment of subjects with MDD, and its effect size is comparable to that of other approved antidepressants.

8.1.2 FKGBE007

Study Description

This was a 9-center, randomized, double-blind, placebo-controlled study evaluating gepirone ER 20-80 mg/day over an 8-week treatment period in subjects with MDD. The study had one forced-titration step from 20-40 mg/day at Day 4, and a flexible-dose design thereafter. The mean dose (± SD) of gepirone ER was 58.2 (±13.95) mg/day, with 87.9% of subjects reaching a final dose of 60-80 mg/day.

Study Design

Randomized, double-blind, placebo-controlled, clinical study evaluating gepirone ER in the treatment of MDD. The study was conducted across 9-centers which enrolled patients diagnosed with moderate to severe depression. After a 4-7 day placebo wash-out period, eligible subjects were randomized to receive either placebo or gepirone ER tablets (20 to 80 mg) once per day in the morning for 8 weeks (56 days). Efficacy evaluations (HAMD, CGI, and MADRS) were performed at baseline (Day 0) and Weeks 2, 3, 4, 6 and 8.

Entrance Criteria

Inclusion Criteria

Key inclusion criteria were:

- Patients between 18 and 64 years of age;
- Patients with moderate to severe MDD (Diagnostic and Statistical Manual, Fourth Edition [DSM-IV]);
- Significant daily dysphoria for 4 weeks prior to screening;
- Hamilton Depression Rating Scale (HAMD)-17 total score of ≥ 20 at both screening and baseline assessments.

Exclusion Criteria

Key exclusion criteria were:

- At least 20% decrease on the HAMD-17 total score between screening and baseline;
- Patients who were ≥ 65 years, diagnosed with, or taking any medication for any chronic illness (other than psychiatric disorders);
- DSM-IV Axis I diagnosis other than MDD;
- DSM-IV Axis II disorder;
- History of treatment-refractory major depressive episodes defined as incomplete or no therapeutic response to 2 prior courses of at least 1 month duration;
- Current requirement for psychotherapy.

Efficacy Assessments

Primary Efficacy Measure

The primary efficacy measure was the change from baseline on the HAMD-17 total score at study endpoint (week 8 or last visit) for the ITT patient population.

Secondary Efficacy Measures

Key secondary efficacy measures, were as follows:

- Change from baseline on:
    - HAMD-21, HAMD-25, and HAMD-28 total scores;
    - Montgomery-Åsberg Depression Rating Scale (MADRS) total score;
    - HAMD item [1] Depressed Mood;
    - Clinical Global Impression of Severity (CGI-S);
- Responder and Remitter analyses, conducted as follows:
    - Proportion of HAMD-17 responders (≥ 50% change from baseline on the HAMD-17 total score);
    - Proportion of HAMD-25 responders (≥ 50% change from baseline on the HAMD-25 total score,) for patients who satisfied DSM-IV criteria for MDD with atypical depression;
    - Proportion of MADRS responders (≥ 50% change from baseline on the MADRS total score);
    - Proportion of Clinical Global Impression (CGI) responders ("much" or "very much" improved, according to the CGI Global Impression-Improvement [CGI-I] assessment) ;
    - Proportion of remitters (HAMD-17 total score ≤ 7 on a post-baseline assessment).

Statistical Methods

For all continuous variables, ANOVA was carried out with treatment and center as factors. For categorical variables, analyses were performed using the CMH test, adjusting for center. The primary analysis was based on the ITT population using the LOCF procedure, where the ITT population included all randomized subjects who took study drug and had at least one post-baseline efficacy assessment. With LOCF, the last assessment for subjects who dropped out or missed visits was imputed for subsequent visits.

Efficacy analyses were also performed using an Observed Case (OC) procedure with no imputation for missing data. This summary focuses on ITT/LOCF results.

Results

Overall, 248 subjects were randomized (124 gepirone ER, 124 placebo).

As shown in *Table 13*, the drop-out rate was similar in the gepirone ER groups and placebo groups (21.8% vs. 17.7%); a few subjects discontinued gepirone ER due to adverse events (4.0% vs. 2.4%), lack of efficacy (3.2% vs. 2.4%), and other reasons (14.5% vs. 12.9%) such as lost to follow-up and withdrew consent.

Table 13: Subjects Discontinued by Reason in Study FKGBE007

Number of Subjects		Treatment Group		
		gepirone-ER	Placebo	Total
Randomized		124	124	248
Treated		124	124	248
Discontinued	Total	27 (21.8%)	22 (17.8%)	49 (39.5%)
	Adverse events	5 (4.0%)	3 (2.4%)	8 (3.2%)
	Lack of efficacy	4 (3.2%)	3 (2.4%)	7 (2.8%)
	Other reason	18 (14.5%)	16 (12.9%)	34 (27.4%)
Completed Treatment		97	102	199

[*Source: CSR FK-GBE-007 Table 9*]

The results of this study showed statistically significant differences in favor of gepirone ER for the primary efficacy variable (change from baseline HAMD-17 total score using LOCF), and nearly all the secondary efficacy variables.

Table 14 presents the change from baseline in HAMD-17 total score by visit and treatment group for the ITT population (LOCF).

Based on the ITT/LOCF population, the treatment effect was statistically significant for the change from baseline HAMD-17 total score between the gepirone ER and placebo treatment groups at Week 4 (p=0.004), Week 6 (p=0.006) and at Week 8/Endpoint (p=0.032), with an effect size of 2.26 points in favor of gepirone ER at the final visit. The difference was also marginally significant at Week 3 (p=0.081). Results based on the ITT/Observed Cases (OC) analysis were consistent with these findings.

Table 14: HAMD-17 Total Score: Change from Baseline at Each Visit (ITT/LOCF) in Study FKGBE007

Treatment Group		Baseline	LS Mean Change from Baseline				
			Week 2	Week 3	Week 4	Week 6	Week 8/ET
gepirone-ER	n	116	112	116	116	116	116
	Mean	23.9	-4.86	-6.72	-8.69	-9.92	-10.22
	SE	0.25	0.43	0.52	0.57	0.64	0.75
Placebo	N	122	120	122	122	122	122
	Mean	24.2	-4.19	-5.45	-6.37	-7.44	-7.96
	SE	0.27	0.41	0.51	0.56	0.62	0.73
Difference: (gep-ER – placebo)			-0.67	-1.28	-2.32	-2.48	-2.26
p-value			0.265	0.081	0.004	0.006	0.032

Analysis is based on ANOVA with effects for treatment and center; ET=End of Treatment; LS=Least Squares; SE=Standard Error of the Mean
[*Source: CSR FK-GBE-007 Table 15*]

Table 15 presents the results for HAMD-17 at final visit (ITT/LOCF) for each center and overall based on an ANCOVA model, with factors for treatment, center, and baseline value as covariate. This approach was used as a secondary analysis. Covariate adjustment was not planned for the primary analysis but was

included in later analyses undertaken to explore the treatment-by-center interaction.

Overall, the average reduction in HAMD-17 was significantly greater in gepirone-treated subjects than in the placebo group (-10.24 vs -7.79), with an effect size of 2.45 (p=0.018). The treatment by center interaction was significant (p=0.092), indicating some inconsistency in results among centers. Center-specific results for HAMD-17 are displayed in Table 15.

Reductions in HAMD-17 were greater in the gepirone ER group compared to the placebo group in 5 of 8 centers; the treatment effect achieved statistical significance in 2 of these centers (centers 701 and 706, with p=0.005 and p=0.044, respectively). Trends favored placebo in 3 centers: center 999 (pooled centers 703 and 707, that each enrolled fewer than 16 subjects), center 704 and center 705 (each with marked response to placebo); none of these differences achieved statistical significance individually.

This covariate-adjusted analysis confirmed the presence of a statistically significant treatment effect favoring gepirone ER for the primary efficacy variable. Center-specific results indicated that positive and significant findings were predominant at the larger centers with reasonable levels of placebo response (HAMD-17 reductions of less than 8).

Table 15: HAMD-17 Total Score: Change from Baseline at End of Treatment, by Center (ITT/LOCF) in Study FKGBE007

Center	Number of Subjects		Adjusted Mean Change		Treatment Difference		
	gepirone-ER	Placebo	gepirone-ER	Placebo	gepirone-ER – Placebo (95% CI)	SE	p-value
Center 701	21	23	-11.49	-5.34	-6.15 (-10.33, -1.97)	2.07	**0.005**
Center 702	14	16	-12.33	-8.02	-4.31 (-11.30, 2.67)	3.40	0.216
Center 704	22	20	-10.38	-11.63	1.25 (-2.63, 5.13)	1.92	0.519
Center 705	14	15	-5.98	-8.35	2.36 (-4.88, 9.61)	3.53	0.509
Center 706	10	9	-12.89	-5.34	-7.55 (-14.85, -0.24)	3.45	**0.044**
Center 708	12	10	-10.91	-10.51	-0.40 (-6.85, 6.05)	3.08	0.898
Center 709	8	11	-11.55	-4.24	-7.31 (-15.41, 0.79)	3.82	**0.074**
Center 999	15	18	-7.20	-7.94	0.74 (-5.28, 6.75)	2.95	0.804
All Centers	116	122	-10.24	-7.79	-2.45 (-4.47, -0.43)	1.02	**0.018**
Treatment by Center Interaction p-value = **0.092**							

Analysis is based on ANCOVA with terms for treatment, center, and baseline value as a covariate; Center 999 is a pooled center that combines centers 703 and 707
[Source: ISE Table 16, Appendix A Statistical Table 3.2]

An overview of results for all secondary efficacy variables at Endpoint (Week8/End of Treatment) is presented in Table 16. Treatment effects consistently favored gepirone ER over placebo for change from baseline values of secondary efficacy variables at Week 4 through Week 8/ET. At Week 8/End of

Treatment, statistically significant differences between treatment groups were evident for the change from baseline in MADRS (p=0.008; effect size of 3.78), HAMD-21 (p=0.043; effect size of 2.28), HAMD-25 (p=0.029; effect size of 2.80 points), and HAMD-28 (p=0.032; effect size of 3.21). The treatment effect was not statistically significant for the change baseline in HAMD-Item 1 at Week 8/ET (p = 0.101), although the difference of 0.38 points did favor gepirone ER.

Table 16: Summary of Secondary Efficacy Results at Endpoint (ITT/LOCF) in Study FKGBE007

Parameter	Week8/End of Treatment		Difference	p-value
	gepirone-ER	Placebo		
HAMD-17 responders (%)	45.7%	29.5%	16.2%	**0.014**
HAMD-17 Remitters (%)	34.5%	20.5%	14%	**0.019**
HAMD-21 CFB	-11.07	-8.79	-2.28	**0.043**
HAMD-21 responders (%)	46.6%	32.0%	14.6%	**0.031**
HAMD-25 CFB	-12.65	-9.85	-2.80	**0.029**
HAMD-25 responders (%)	48.3%	30.3%	18.0%	**0.007**
HAMD-28 CFB	-15.04	-11.83	-3.21	**0.032**
HAMD-28 responders (%)	49.1%	32.8%	16.3%	**0.015**
CGI severity CFB	-1.30	-0.92	-0.38	**0.015**
CGI responders (%)	48.3%	34.7%	13.6%	**0.045**
HAMD Item 1 CFB	-1.22	-0.97	-0.32	0.101
MADRS CFB	-13.72	-9.94	-3.78	**0.008**
MADRS responders (%)	50.9%	32.2%	18.7%	**0.005**

[*Source: FK-GBE-007 CSR Table 22, Appendix 15 Supportive Tables 22, 29, 30, 43, 44, 51, 52, 59, 60, 67, 73,74, 103*]
CFB=Change from baseline; LS means are from ANOVA model, including effects for treatment and center; % responders and % remitters were compared using CMH test. For HAMD and MADRS scales, responders are subjects with ≥ 50% reduction from baseline at any post-baseline assessment; HAMD-17 Remitters are subjects with a HAMD-17 total score of ≤ 7. For CGI, responders are much or very much improved on the CGI improvement score at any post-baseline assessment.

Responder rates are also shown in Table 16. The number (%) of responders based on the HAMD, MADRS, and CGI scales were consistently higher in the gepirone ER group than the placebo group from Week 3 through Week 8/End of Treatment. At Week 8/End of Treatment, statistically significant differences between treatment groups were detected in the number (%) of HAMD-17 responders (p=0.014; 16.2% in favor of gepirone ER), MADRS (p=0.005; 18.7% in favor of gepirone ER), HAMD-25 responders (p=0.007; 18.0% difference in favor of gepirone ER), HAMD-28 responders (p=0.015; 16.3% difference in favor of gepirone ER), CGI Responders (p=0.045; 13.6% difference in favor of gepirone ER), and HAMD-17 remitters (p=0.019; 14.0% difference in favor of gepirone ER) based on the LOCF analysis for the ITT group. Results based on the LOCF and OC analyses for the ITT Group were consistent.

Conclusions

This study was a 9-center, randomized, double-blind, placebo-controlled study, which employed doses in the appropriate therapeutic range for gepirone ER. Results for the primary efficacy variable were statistically significant and supported by secondary efficacy variables.

The treatment effect was statistically significant favoring gepirone ER over placebo for the primary efficacy variable, and this positive result was supported by nearly all secondary efficacy variables.

Study FKGBE007 provides strong evidence of the therapeutic effectiveness of gepirone ER for the treatment of subjects with MDD, and its effect size is comparable to that of other approved antidepressants.

8.1.3 134002

Study Description

This was a 5-center, randomized, double-blind, placebo-controlled study evaluating gepirone ER 20-80 mg/day over an 8-week treatment period in subjects with MDD. The study had a flexible-dose design, and the minimum final dose of gepirone ER was 40 mg/day. The mean dose (± SD) of gepirone ER was 57.90 (± 13.03) mg/day. The final prescribed dose was 60 and 80 mg/day in 23.4% and 58.9% of subjects, respectively.

The primary objective of this study was to evaluate the therapeutic efficacy of gepirone ER in comparison with placebo at the endpoint of an 8 week treatment period in subjects with major MDD, diagnosed according to DSM-IV criteria.

Study Design

Randomized, double-blind, placebo-controlled, clinical study evaluating gepirone ER in the treatment of MDD. The study was conducted across 5-centers which enrolled patients diagnosed with moderate to severe depression and a HAMD-17 total screen and baseline score of > 20.

After a 7 day placebo wash-out period, eligible subjects were randomized to receive either placebo or gepirone ER tablets (20 to 80 mg) once per day in the morning for 8 weeks (56 days). Efficacy evaluations (HAMD, CGI, and MADRS) were performed at baseline (Day 0) and Weeks 1, 2, 3, 4, 6 and 8.

Diagnosis and Inclusion Criteria

Eligible subjects were 18-70 years of age, met diagnostic criteria for moderate to severe MDD according to DSM-IV criteria, had a HAMD-17 total score of ≥20 at both screening and baseline assessments, had significant daily dysphoria for the past four weeks, and provided written informed consent.

Efficacy Assessments

Primary Efficacy Measure

The primary efficacy parameter was the change from baseline in HAMD-17 total score at endpoint (Week 8 or last visit) in the ITT population using the LOCF procedure. The ITT group included all randomized subjects who took study drug and had at least one post-baseline efficacy assessment within 3 days after the last dose of study drug. With LOCF, the last assessment for subjects who discontinued treatment was imputed for all subsequent visits.

Secondary Efficacy Measures

- Change from baseline on:
    - HAMD-21, HAMD-25, and HAMD-28 total scores;
    - Montgomery-Åsberg Depression Rating Scale (MADRS) total score;
    - HAMD item [1] Depressed Mood;
    - HAMD anxiety/somatization factor (Factor 1);
    - HAMD item 12 (psychic anxiety);
    - Clinical Global Impression of Severity (CGI-S);
    - Bech 6;

- Responder and Remitter analyses, conducted as follows:
    - Proportion of HAMD-17 responders (≥ 50% change from baseline on the HAMD-17 total score);
    - Proportion of HAMD-25 responders (≥ 50% change from baseline on the HAMD-25 total score,) for patients who satisfied DSM-IV criteria for MDD with atypical depression;
    - Proportion of Clinical Global Impression (CGI) responders ("much" or "very much" improved, according to the CGI Global Impression-Improvement [CGI-I] assessment) ;
    - Proportion of remitters (HAMD-17 total score ≤ 7 on a post-baseline assessment).

Statistical Methods

For all continuous variables, ANOVA was carried out with treatment and center as factors. For categorical variables, analyses were performed using the CMH test, adjusting for center.

Note: FKP performed post-hoc mixed model analyses for HAMD-17 and several other parameters, including the modified HAMD-17 (mHAMD-17), the HAMD core depression factor, HAMD Item 1, and the MADRS total score. Results of the mixed-models analyses are presented under 2 different covariance structures:

first-order autoregressive structure (Model 1) and compound symmetry (Model 2) for the ITT dataset.

Results

Overall, 219 subjects were randomized and 218 were treated (110 gepirone ER, 108 placebo). A total of 211 subjects with post-baseline data comprised the ITT population (107 gepirone ER, 104 placebo). Six subjects had post-baseline efficacy assessment more than 3 days after the last dose of study medication (1 placebo, 5 gepirone ER) and were not included in any analyses.

As shown in *Table 17*, discontinuation rates were comparable in gepirone ER and placebo groups (31.8% and 28.7%, respectively). The majority of drop-outs were for unspecified reasons (19.1% vs. 18.5%) such as lost to follow-up and withdrawn consent. Three subjects in each group dropped out for lack of efficacy.

Table 17: Subjects Discontinued by Reason in Study 134002

Number of Subjects		Treatment Group		
		gepirone-ER	Placebo	Total
Randomized		110	109	219
Treated		110	108	218
Discontinued	Total	35 (31.8%)	31 (28.7%)	66 (30.3%)
	Adverse events	11 (10.0%)	8 (7.4%)	19 (8.7%)
	Lack of efficacy	3 (2.7%)	3 (2.8%)	6 (2.8%)
	Other reason	21 (19.1%)	20 (18.5%)	41 (18.8%)
Completed Treatment		75	77	152

†One randomized subject returned all study medication and was excluded from ITT.
Source: CSR Table 5, Appendices F8.1.1-1 and F8.1.2-1

The primary efficacy variable (change from baseline in HAMD-17) showed greater improvement in the gepirone ER group compared to placebo, but the differences did not achieve statistical significance at endpoint or any other study visit. Trends in mean values directionally favored gepirone ER over placebo consistently from Week 1 to endpoint. *Table 18* provides the change from baseline in HAMD-17 total score by visit and treatment group for the ITT population (LOCF).

Table 18: HAMD-17 Total Score: Change from Baseline at Each Visit (ITT/LOCF) in Study 134002

Treatment Group		Baseline	LS Mean Change from Baseline					
			Week 1	Week 2	Week 3	Week 4	Week 6	Week 8/ET
gepirone-ER	n	107	101	102	102	102	102	102
	Mean	23.96	-4.34	-6.68	-8.71	-9.50	-9.92	-9.96
	SE	0.27	0.38	0.46	0.55	0.63	0.67	0.65
Placebo	n	104	101	103	103	103	103	103
	Mean	24.11	-3.73	-6.12	-8.04	-8.66	-9.74	-9.29
	SE	0.29	0.38	0.46	0.55	0.63	0.67	0.65
Difference: (gep-ER – placebo)			-0.61	-0.56	-0.67	-0.84	-0.18	-0.67
p-value			0.235	0.375	0.370	0.322	0.841	0.446

Based on ANOVA with effects for treatment and center; ET=End of Treatment; LS=Least Squares; SE=Standard Error of the Mean
[*Source: CSR 134002 Table 12, Appendices F8.6.1.1-2 and F8.6.1.1-4*]

An overview of results for all secondary efficacy variables at Endpoint (Week 8/End of Treatment) is presented in *Table 19*. Data for some additional time points are discussed, but not shown.

Overall, treatment effects consistently favored gepirone ER over placebo for each of the secondary efficacy variables at all study visits.

Differences achieved statistical significance for HAMD Item 1 (depressed mood) at Week 2 (p=0.027), Week 3 (p=0.015), Week 6 (p=0.040) and Week 8/ET (p=0.036). Trends favored gepirone ER for responder rates, but the differences did not achieve statistical significance at the 5% level. The number (%) of CGI Responders was higher in the gepirone ER group compared to placebo at Week 4 (p=0.064) and Week 6 (p=0.057). By Week 8/Endpoint, the difference was not statistically significant.

For the number (%) of HAMD-25 responders, there were also marginally significant differences between treatments favoring gepirone ER at Week 3 (p=0.079), Week 4 (p=0.090), and Week 8/ET (p = 0.052; difference of 12.2% in responder rates). Similar results were seen in the ITT/OC analysis, except a statistically significant difference in the number (%) of HAMD-17 responders was noted at Visit 6 (p = 0.044) and in the number (%) of HAMD-25 responders at Week 8/ET (p =0.011).

Table 19: Summary of Secondary Efficacy Results at Endpoint (ITT/LOCF) in Study 134002

Parameter	End-of-Treatment Outcome		Difference	p-Value
	gepirone-ER	Placebo		
HAMD-17 Responders (%)	40.2%	32.0%	8.2%	0.225
HAMD-17 Remitters (%)	15.7%	17.5%	-1.8%	0.731
HAMD-21 CFB	-10.77	-9.90	-0.87	0.365
HAMD-25 CFB	-12.54	-11.00	-1.54	0.153
HAMD-25 Responders (%)	42.2%	29.1%	13.1%	0.014
HAMD-28 CFB	-14.32	-12.51	-1.81	0.150
HAMD-Item 1 CFB	-1.30	-1.01	-0.18	0.036
HAMD-Factor 1 CFB	-3.14	-2.77	0.37	0.277
CGI Severity CFB	-1.09	-0.88	-0.21	0.130
CGI Improvement	2.51	2.73	-0.22	0.145
CGI Responders (%)	52.0%	44.7%	7.3%	0.297
MADRS CFB	-11.55	-9.21	-2.34	0.078
Bech-6 CFB	-5.90	-4.95	-0.95	0.076

[*Source: 134002 CSR Table 21, Appendix F8.6.2*]
CFB=Change from baseline; LS means from ANOVA model, including effects for treatment and center; % responders and % remitters were compared using CMH test. CGI Responders are subjects much or very much improved based on the CGI improvement score at a post-baseline assessment. HAMD-17 Responders are subjects with ≥ 50% reduction from baseline at a post-baseline assessment. HAMD-25 Responders are subjects with ≥ 50% reduction from baseline at the endpoint assessment.

To further evaluate positive efficacy trends noted above, a post-hoc mixed models ANOVA model was applied to HAMD-17 and several other parameters,

including the modified HAMD-17 (mHAMD-17)[3], the HAMD core depression factor (Bech-6), HAMD Item 1, and the MADRS total score.

These data are presented in Table 20. The results of the mixed-models analyses for two different mixed models approaches, demonstrated statistically significant treatment effects favoring gepirone ER over placebo for each of the secondary efficacy variables. Results are strongly positive for the mHAMD-17, Bech-6, Item-1, and MADRS in both analyses.

Table 20: Post-Hoc Mixed Models Analyses Results in Study 134002

Parameter	Change from Baseline		Mixed Models p-value
	gepirone-ER	Placebo	
HAMD-17			
Mixed Model 1	-8.81 ± 0.42	-7.97 ± 0.41	0.135
Mixed Model 2	-8.93 ± 0.42	-8.08 ± 0.41	0.137
mHAMD-17			
Mixed Model 1	-8.27 ± 0.41	-6.67 ± 0.51	0.004
Mixed Model 2	-8.33 ± 0.41	-6.75 ± 0.41	0.005
Bech-6			
Mixed Model 1	-5.25 ± 0.26	-4.06 ± 0.25	0.001
Mixed Model 2	-5.30 ± 0.26	-4.10 ± .025	0.001
HAMD Item 1			
Mixed Model 1	-1.20 ± 0.07	-0.89 ± 0.06	0.001
Mixed Model 2	-1.21 ± 0.07	-0.90 ± 0.06	0.001
MADRS			
Mixed Model 1	-10.63 ± 0.61	-8.09 ± 0.60	0.002
Mixed Model 2	-10.71 ± 0.61	-8.18 ± 0.60	0.002
[*Source: ISE Table 20*]			

Conclusions

This study was a 5-center, randomized, double-blind, placebo-controlled study, which employed doses in the appropriate therapeutic range for gepirone ER.

While results for the primary efficacy variable were not statistically significant, positive trends were noted for secondary efficacy variables and highly significant treatment effects were obtained using a post-hoc repeated measures (mixed models) analysis, providing supportive evidence that gepirone ER has anti-depressant activity.

[3] Among the secondary outcome measures, the mHAMD-17 is a modification that replaces 5 items related to insomnia and appetite (Insomnia early, Insomnia middle, Insomnia late, Somatic symptoms gastrointestinal, and Loss of weight) with HAMD-25 items that measure the opposite (or reverse) neuro-vegetative symptoms (Hypersomnia: Time in bed, Oversleeping, and Napping; Increased appetite, Weight gain). These item substitutions provide a scale that removes the effect of 5 items that may be less sensitive in some subjects who are receiving compounds that can produce insomnia, nausea, and agitation, but that includes relevant reverse neurovegetative symptoms frequent in MDD.

8.1.4 FKGBE008

Study Description

This was an 8-center randomized, double-blind, placebo-controlled, flexible dose study evaluating gepirone ER 20-80 mg/day over an 8-week treatment period in subjects with MDD. The mean dose of gepirone ER was 60.0±13.1 mg/day. By the final visit, 86.9% of subjects were at a dose of 60-80 mg/day.

The primary objective of this study was to evaluate the therapeutic efficacy of gepirone ER in comparison with placebo at the end of an 8 week treatment period in subjects with MDD.

Study Design

The study design is the same as Study FKGBE007 (See Section 3.3.2), except for a smaller planned sample size: 100 subjects/group (instead of 120 subjects/group in FKGBE007).

Diagnosis and Inclusion Criteria

Eligible subjects were males or females, 18-64 years of age who met DSM-IV criteria for moderate to severe MDD, had a HAMD-17 score of ≥20 at screening and baseline, and had significant daily dysphoria for 4 weeks prior to screening.

8.1.4.1 Efficacy Assessments

8.1.4.1.1 Primary Efficacy Measure

The primary efficacy measure was the change from baseline on the HAMD-17 total score at study endpoint for the Intention-to-Treat (ITT) patient population.

8.1.4.1.2 Secondary Efficacy Measures

Key secondary efficacy measures, were as follows:

- Change from baseline on:
    - HAMD-21, HAMD-25, and HAMD-28 total scores;
    - Montgomery-Åsberg Depression Rating Scale (MADRS) total score;
    - HAMD item [1] Depressed Mood;
    - HAMD anxiety/somatization factor (Factor 1);
    - Clinical Global Impression of Severity (CGI-S);

- Responder and Remitter analyses, conducted as follows:

- Proportion of HAMD-17 responders (≥ 50% change from baseline on the HAMD-17 total score);
- Proportion of HAMD-25 responders (≥ 50% change from baseline on the HAMD-25 total score,) for patients who satisfied DSM-IV criteria for MDD with atypical depression;
- Proportion of MADRS responders (≥ 50% change from baseline on the MADRS total score);
- Proportion of remitters (HAMD-17 total score ≤ 7 on a post-baseline assessment).

8.1.4.2 Statistical Methods

For all continuous variables, ANOVA was carried out with treatment and center as factors. For categorical variables, analyses were performed using the CMH test, adjusting for center. The primary analysis was based on the ITT population using the LOCF procedure. The ITT population included all randomized subjects who took study drug and had at least one post-baseline efficacy assessment. With LOCF, the last assessment for subjects who dropped out or missed visits was imputed for subsequent visits. Efficacy analyses were also performed using the OC procedure (no imputation for missing data). This summary focuses on ITT/LOCF results.

8.1.4.3 Results

Overall, 206 subjects were randomized and received at least one dose of study medication (102 gepirone ER, 104 placebo group). A total of 199 subjects comprise the ITT population (99 gepirone ER, 100 placebo).

As shown in *Table 21*, the drop-out rate was similar in the gepirone ER groups and placebo groups (24.5% vs. 21.1%). A few more subjects discontinued gepirone ER due to adverse events (4.9% vs. 1.9%), but drop-outs for lack of efficacy or other reasons (such as withdrawn consent or non-compliance) occurred with low and comparable frequency in the two groups.

Table 21: Subjects Discontinued by Reason in Study FKGBE008

Number of Subjects		Treatment Group		
		gepirone-ER	Placebo	Total
Randomized		102	104	206
Treated		102	104	206
Discontinued	Total	25 (24.0%)	22 (21.5%)	47 (22.8%)
	Adverse events	5 (4.9%)	2 (1.9%)	7 (3.4%)
	Lack of efficacy	1 (1.0%)	2 (1.9%)	3 (1.5%)
	Other reason	19 (18.2%)	18 (17.6%)	47 (22.8%)
Completed Treatment		77	82	159

[Source: CSR FK-GBE-008 Table 10]

Mean reductions in HAMD-17 were greater in the gepirone ER group than in the placebo group at each visit; differences were statistically significant at Week 2 and Week 6. However, by the end of the study (Week 8), the difference between groups was no longer statistically significant. Data are presented in *Table 22* as the change from baseline in HAMD-17 total score at each visit for the ITT population (LOCF).

Results based on the LOCF and OC analyses for the Intent-to-Treat Group were consistent.

Table 22: HAMD-17 Total Score: Change from Baseline at Each Visit (ITT/LOCF) in Study FKGBE008

Treatment		Baseline	LS Mean Change from Baseline				
			Week 2	Week 3	Week 4	Week 6	Week 8/ET
gepirone-ER	n	99	93	96	96	96	96
	Mean	24.2	-5.62	-7.31	-8.52	-9.46	-9.87
	SE	0.30	0.44	0.55	0.62	0.69	0.75
Placebo	n	100	93	98	98	98	98
	Mean	24.0	-4.10	-5.81	-7.17	-7.52	-8.37
	SE	0.27	0.44	0.54	0.61	0.68	0.75
Difference: (gep-ER – placebo)			-1.52	-1.51	-1.35	-1.94	-1.50
p-value			0.016	0.053	0.123	0.046	0.159

ET=End of treatment; LS=Least squares; SE=Standard error
[*Source: CSR FK-GBE-008 Table 16, Supportive Table 21*]

P-values resulting from statistical tests of treatment effect for secondary efficacy variables at each study visit and endpoint are presented in Table 23.

Trends in mean change values consistently favored gepirone ER over placebo for each of the secondary efficacy variables at all study visits. Differences achieved statistical significance for MADRS (Weeks 2, 3 and 4) and for HAMD-21, HAMD-28, and HAMD-Item 1 at Week 2. Though trends were positive, there were no statistically significant differences at Week 8/End of Treatment.

The number (%) of responders based on all efficacy scales were consistently higher in the gepirone ER group than the placebo group at each visit, and differences achieved statistical significance in a few instances for HAMD-17, CGI, and HAMD-25. At Week 8/End of Treatment, a statistically significant difference between treatment groups was detected in the number (%) of HAMD-25 responders (p=0.035; 14.2% in favor of gepirone ER).

Results based on the LOCF and OC analyses for the ITT Group were consistent.

Table 23: Statistical Significance of Secondary Efficacy Results at Each Time Point (ITT/LOCF) in Study FKGBE008

Parameter	Week 2	Week 3	Week 4	Week 6	Week 8/ET
HAMD-17 Responders	0.096	0.176	0.287	0.050	0.293
MADRS CFB	0.008	0.039	0.035	0.098	0.208
MADRS Responders	0.159	0.310	0.271	0.220	0.128
HAMD-21 CFB	0.023	0.147	0.170	0.055	0.209
HAMD-21 Responders	0.054	0.552	0.458	0.068	0.231
HAMD-25 CFB	0.057	0.135	0.243	0.087	0.281
HAMD-25 Responders	0.288	0.676	0.215	0.024	0.035
HAMD-28 CFB	0.025	0.212	0.308	0.119	0.319
HAMD-28 Responders	0.259	0.556	0.311	0.104	0.178
HAMD Item 1 CFB	0.035	0.321	0.649	0.552	0.469
CGI Severity	0.101	0.214	0.075	0.070	0.273
CGI Responders	0.132	0.011	0.100	0.037	0.147
HAMD-17 Remitters	0.162	0.756	0.448	0.122	0.156

[*Source: FK-GBE-008 CSR Appendix 15 Supportive Tables 22, 29-30, 43-44, 51-52, 59-60, 67, 73-74 and 103*]
P-values are from ANOVA model, including effects for treatment and center; % responders and % remitters were compared using CMH test.
CFB=Change from baseline. For HAMD and MADRS scales, responders are subjects with ≥ 50% reduction from baseline at any post-baseline assessment; HAMD-17 Remitters are subjects with a HAMD-17 total score of ≤ 7. For CGI, responders are much or very much improved on the CGI improvement score at any post-baseline assessment.

Table 24 presents an overview of results for secondary efficacy variables at the end of treatment (8 weeks).

Subjects treated with gepirone ER experienced numerically greater symptom improvement compared to placebo-treated subjects throughout the study. However, by Week 8, the differences between treatments were not statistically significant at the 5% significance level.

Responder rates for the HAMD-17, HAMD-21, HAMD-25, HAMD-28, CGI, and MADRS scales were consistently higher for the gepirone ER group compared to placebo. The HAMD-25 responder rate was significantly greater in the gepirone ER group at endpoint; however, responder rates for the other scales did not achieve significance at Week 8.

Likewise, the number of subjects classified as HAMD-17 remitters (subjects with post-baseline HAMD-17 score of ≤ 7) was numerically greater in the gepirone ER group than in the placebo group at all time points, but the differences failed to reach statistical significance.

Table 24: Summary of Secondary Efficacy Results at Endpoint (ITT/LOCF) in Study FKGBE008

Parameter	Week8/End of Treatment		Difference	p-value
	gepirone-ER	Placebo		
HAMD-17 responders (%)	39.6%	32.7%	6.9%	0.293
MADRS CFB	-11.73	-9.87	-1.86	0.208
MADRS responders (%)	39.6%	29.6%	10.0%	0.128
HAMD-21 CFB	-10.67	-9.14	-1.52	0.209
HAMD-21 responders (%)	40.6%	32.7%	7.9%	0.231
HAMD-25 CFB	-11.51	-10.14	1.37	0.281
HAMD-25 responders (%)	44.8%	30.6%	14.2%	0.035
HAMD-28 CFB	-13.47	-11.94	-1.53	0.319
HAMD-28 responders (%)	42.7%	33.7%	9.0%	0.178
HAMD Item 1 CFB	-1.10	-0.99	-0.11	0.469
CGI severity CFB	-1.27	-1.08	-0.19	0.275
CGI responders (%)	47.9%	37.8%	10.1%	0.147
HAMD-17 remitters (%)	22.9%	15.3%	7.6%	0.156

[*Source: FK-GBE-008 CSR Appendix 15 Supportive Tables 22, 29-30, 43-44, 51-52, 59-60, 67, 73-74 and 103*]
CFB=Change from baseline; LS means are from ANOVA model, including effects for treatment and center; % responders and % remitters were compared using CMH test. For HAMD and MADRS scales, responders are subjects with ≥ 50% reduction from baseline at any post-baseline assessment; HAMD-17 remitters are subjects with a HAMD-17 total score of ≤ 7. For CGI, responders are much or very much improved on the CGI improvement score at any post-baseline assessment.

8.1.4.4 Conclusions

Overall, while this study failed to demonstrate a statistically significant treatment effect for the protocol-defined primary endpoint (change from baseline in HAMD-17 scores at Week 8), trends in mean values directionally favored gepirone ER over placebo at each visit, with significant differences detected at Week 2 and Week 6 (p=0.016 and p=0.046, respectively). By Week 8, the mean change from baseline scores were -9.9 and -8.4 (p=0.159).

These data suggest that gepirone ER has a beneficial impact on symptoms of depression.

8.1.5 134023

Study Description

This was a 12-center, randomized, double-blind, placebo-controlled, flexible dose trial evaluating gepirone ER in subjects with MDD during a 9-week treatment period. The mean dose (± SD) of gepirone ER was 61.3 (±13.66) mg/day, with 69.3% of subjects reaching a final dose of 80 mg/day.

The primary objective of this study was to evaluate the therapeutic efficacy of gepirone ER in comparison with placebo at the endpoint of a 9-week treatment period in subjects with MDD.

Study Design

Randomized, double-blind, placebo-controlled, clinical study evaluating gepirone ER in the treatment of MDD. The study was conducted across 12-centers which enrolled patients diagnosed with moderate to severe depression.

After a 4-14 day placebo wash-out period, eligible subjects were randomized to receive either placebo or gepirone ER tablets (20 to 80 mg) once per day in the morning for 9 weeks (63 days). Subjects were evaluated at baseline (Day 0) and Weeks 1, 2, 3, 5, 7 and 9.

Diagnosis and Inclusion Criteria

Subjects were 18-70 years of age with moderate to severe MDD (diagnosed according to DSM-IV criteria), had experienced the current episode of MDD for a minimum of 1 month, had a MADRS total score ≥30 at screening and baseline, and had dysphoria for most days over the past 4 weeks.

Efficacy Assessments

Primary Efficacy Measure

The primary efficacy parameter was the change from baseline in HAMD-17 total score at endpoint (Week 9 or last visit) in the ITT group using the LOCF procedure.

Secondary Efficacy Measures

- Change from baseline on:
    - HAMD-25 total score;
    - Montgomery-Åsberg Depression Rating Scale (MADRS) total score;
    - HAMD item [1] Depressed Mood;
    - Clinical Global Impression of Severity (CGI-S);
    - CGI Improvement Score (CGI-I);
    - Bech-6;

- Responder and Remitter analyses, conducted as follows:
    - Proportion of HAMD-17 responders (≥ 50% change from baseline on the HAMD-17 total score);
    - Proportion of CGI responders;
    - Proportion of remitters (HAMD-17 total score ≤ 7 on a post-baseline assessment).

Statistical Methods

For all continuous variables, analysis of variance (ANOVA) was carried out with treatment and center as factors. For categorical variables, analyses were performed using the CMH test, adjusting for center.

The ITT group included all randomized subjects who took study drug and had at least one post-baseline efficacy assessment. With LOCF, the last assessment for subjects who discontinued treatment was imputed for all subsequent visits.

Results

Overall, 255 subjects were randomized, and 254 received study treatment (127 per group). A total of 246 subjects (123 per group) with post-baseline data comprised the ITT population.

As shown in *Table 25*, the discontinuation rate was similar in the two treatment groups (26.0% vs. 21.3%), with more subjects discontinuing gepirone ER for adverse events (9.4% vs. 0.1%), but fewer for lack of efficacy (3.9% vs. 5.5%). Most of the drop-outs were due to "other" reasons (12.6% vs. 15.0%), including lost to follow-up, withdrawn consent, and protocol non-compliance.

Table 25: Subjects Discontinued by Reason in Study 134023

Number of Subjects		Treatment Group		
		gepirone-ER	Placebo	Total
Randomized		127	128	255
Treated		127	127	254
Discontinued	Total	33 (26.0%)	27 (21.3%)	60 (23.6%)
	Adverse events	12 (9.4%)	1 (0.1%)	13 (5.1%)
	Lack of efficacy	5 (3.9%)	7 (5.5%)	12 (4.7%)
	Other reason	16 (12.6%)	19 (15.0%)	35 (13.8%)
Completed Treatment		94 (74.0%)	100 (78.7%)	194 (76.4%)

[*Source: CSR ORG1340023 Tables 7 and 8*]

As shown in *Table 26*, there were no statistically significant differences between gepirone ER and placebo groups in the mean change from baseline for HAMD-17 total score at any visit, including Week9/ET (End of Treatment). These results are consistent with other analyses, including the ITT/OC analysis.

Table 26: HAMD-17 Total Score: Change from Baseline at Each Visit (ITT/LOCF) in Study 134023

Treatment Group		Baseline	LS Mean Change from Baseline					
			Week 1	Week 2	Week 3	Week 5	Week 7	Week 9/ET
gepirone-ER	n	123	123	123	123	123	123	123
	Mean	22.9	-2.8	-4.7	-6.2	-7.6	-8.3	-8.0
	SE	0.32	0.34	0.44	0.55	0.61	0.65	0.72
Placebo	n	123	123	123	123	123	123	123
	Mean	22.8	-3.1	-4.9	-6.8	-7.5	-8.5	-8.0
	SE	0.33	0.34	0.45	0.56	0.61	0.66	0.72
Difference: (gep-ER – placebo)			0.3	0.2	0.6	-0.1	0.2	0.0
p-value			0.558	0.815	0.433	0.903	0.829	0.947

[*Source: CSR 134023 Table 13*]

Table 27 summarizes endpoint results (ITT/LOCF) for the secondary efficacy variables. There were no statistically significant differences between treatment groups for any of the secondary efficacy variables. Similar results were obtained in the ITT/OC analysis.

Table 27: Summary of Secondary Efficacy Results at Endpoint (ITT/LOCF) in Study 134023

Parameter	End-of-Treatment Outcome		Difference	p-Value
	gepirone-ER	Placebo		
HAMD-17 Responders (%)	33%	37%	-4%	0.518
HAMD-17 Remitters (%)	22%	24%	-2%	0.731
HAMD-25 CFB	-10.4	-10.0	-0.4	0.739
HAMD-Item 1 CFB	-1.1	-1.2	0.1	0.438
CGI Severity CFB	-1.2	-1.3	0.1	0.467
CGI Improvement	2.7	2.7	0.0	0.864
CGI Responders (%)	43%	39%	4%	0.624
MADRS CFB	-13.9	-13.1	-0.8	0.572
Bech-6 CFB	-4.6	-4.5	-0.1	0.906

[*Source: CSR 134023 Tables 14-22*]
CFB=Change from baseline; LS means are from ANOVA model, including effects for treatment and center; % responders and % remitters were compared using CMH test. For HAMD-17 responders are subjects with ≥ 50% reduction from baseline at any post-baseline assessment; HAMD-17 remitters are subjects with a HAMD-17 total score ≤ 7. For CGI, responders are much or very much improved on the CGI improvement score at any post-baseline assessment.

Conclusions

Overall, no statistically significant treatment effects were detected for gepirone ER based on the primary or secondary efficacy variables. HAMD-17 scores showed an average reduction of 8 points in both treatment groups at the end of study (p=0.947).

While adequately designed, 134023 is a negative study.

9. APPENDIX 2 – SUMMARIES OF 7 UNINTERPRETABLE STUDIES

APPENDIX 2

SUMMARIES OF 7 UNINTERPRETABLE STUDIES

9.1.1 CN105-078

Study Description

This study was a 2-center, randomized, double-blind, placebo-controlled flexible dose trial evaluating gepirone ER at 2 dose levels (10-50 mg/day [low dose] and 20-100 mg/day [high dose]) in subjects with MDD during a 6-week treatment period. The mean maximum dose of gepirone ER administered during this study was 37.3 ±11.2 mg/day in the gepirone ER low-dose group, compared to 67.5±22.0 mg/day in the gepirone ER high-dose group. Thus, half of subjects in the gepirone ER groups received maximum doses below the minimum effective dose (40 mg/day).

Study Design

Randomized, double-blind, placebo-controlled, clinical study evaluating gepirone ER in the treatment of MDD. The study was conducted at 2 sites in the US and was designed to enroll 180 subjects (120 in the combined gepirone ER groups and 60 in the placebo group), which would provide at least 72% power. Study was terminated prematurely when BMS discontinued development of gepirone ER. At the time of termination, 146 subjects were randomized and 144 subjects were treated, thus reducing the power of the study to approximately 62%.

After a 4-day to 4-week baseline period, eligible subjects were randomized to receive either placebo, gepirone ER 10-50 mg/day, or gepirone ER 20-100 mg/day for 6 weeks. Subjects were evaluated at baseline (Day 0) and Weeks 1, 2, 3, 4, and 6 or early termination. Gepirone ER was provided in 10 mg or 20 mg tablets, with matching placebo tablets, administered orally. The daily dosing regimen was 1-5 tablets and individualized for each subject. Double-blind treatment was administered as 1-2 tablets QD for the first week, 3-4 tablets/day (QD or BID) for Weeks 2-3, and 5 tablets/day (QD or BID) at the end of Week 3 depending on response and tolerability. The target dose range for all groups was 3-5 tablets/day.

Entrance Criteria

Inclusion Criteria

Key inclusion criteria were:

- Patients at least 18 years of age;
- Patients with a diagnosis of MDD by DSM-III-R (single episode or recurrent)
- Patients with a baseline HAMD-17 score $\geq$ 20

Efficacy Assessments

Primary Efficacy Measure

The co-primary efficacy parameters were the mean change from baseline in the HAMD-17 total score and the number (%) of CGI responders ("much improved" or "very much improved") at Week 6/End of Treatment based on the ITT/LOCF analysis.

Secondary Efficacy Measures

Key secondary efficacy measures, were as follows:

- Change from baseline on:

  - HAMD-25
  - HAMD-28
  - HAMD-Item 1
  - HAMD-17 responder rate
  - HAMD Factor I (anxiety/somatization)
  - HAMD Factor V (retardation)
  - HAMD Factor VI (sleep disturbance)
  - MADRS
  - CGI-severity score
  - CGI-global improvement

Statistical Methods

The primary efficacy analysis was a comparison of pooled gepirone ER dose groups vs. placebo, whereas the secondary analysis compared each individual gepirone ER dose group with placebo.

The original intent of the protocol was to pool the two sites, if appropriate, with justification provided (e.g., a test of treatment-by-site interaction). While no quantitative interaction was observed, the CSR notes a qualitative difference in placebo response at the two centers, possibly due to differences in clinical profiles of patients at each center. For this reason, results for the primary efficacy parameters were presented separately for each study site and for both sites combined.

For continuous data involving both centers, an ANOVA model was used with treatment and center as factors, and including the treatment-by-center interaction term. For by-center analysis, one-way ANOVA was used with treatment as a factor. CMH tests were used for categorical data, controlling for center. In addition, by-center analyses were presented.

Results

Overall, 144 subjects were randomized and treated (50 geperione ER 10-50mg/day, 45 gepirone ER 20-100 mg/day, 49 placebo)

As shown in Table 28, 55 subjects (38.2%) did not complete the 6-week treatment period. Reasons for drop-out were adverse events (10.0% low dose, 31.1% high dose, and 8.2% placebo), lack of efficacy (2.0% low dose, 0.0% high dose, and 10.2% placebo), and other reasons (14.0% low dose, 20.0% high dose, and 10.2% placebo) that included subject unreliability, lost to follow-up and withdrew consent. The most common reasons for discontinuation were adverse events in the gepirone ER groups and lack of efficacy in the placebo group.

Table 28: Subjects Discontinued by Reason (Combined Sites)

Number of Subjects		Treatment			
		Gepirone ER 10-50 mg/day	**Gepirone ER 20-100 mg/day**	**Placebo**	**Total**
Randomized		50	47	49	146
Treated		50	45	49	144
Discontinued	Total	17 (34.0%)	23 (51.1%)	15 (30.6%)	55 (38.2%)
	Adverse events	5 (10.0%)	14 (31.1%)	4 (8.2%)	23 (16.0%)
	Lack of efficacy	1 (2.0%)	0	5 (10.2%)	6 (4.2%)
	Discontinued by BMS	4 (8.0%)	0	1 (2.0%)	5 (3.5%)
	Other reason	7 (14.0%)	9 (20.0%)	5 (10.2%)	21 (14.6%)
Completed Treatment		33 (66.0%)	22 (48.9%)	34 (69.4%)	89 (61.8%)

[*Source: CSR CN105-078 Tables 11 and 15 and Appendix F Tables 6.5-1 and 6.5-1A*]

Table 29 presents the change from baseline in the HAMD-17 score by visit and treatment group for the ITT population (LOCF).

There were no statistically significant differences between gepirone ER and placebo at any time point for the mean change from baseline for the HAMD-17 total score in the ITT population (LOCF analysis). Differences numerically favored the combined gepirone ER dose group over placebo at End of Treatment for the combined sites and for site 0002.

These results are generally consistent with the other analyses, although the ITT/OC analysis showed statistically significant differences in favor of the high-dose and combined gepirone ER dose groups compared to placebo at Week 4.

Table 29: HAMD-17 Total Score: Change from Baseline at Each Visit (ITT/LOCF) — Study CN105-078

Treatment		Baseline	LS Mean Change from Baseline				
			Week 1	Week 2	Week 3	Week 4	Week 6/ET
Site 001 (Cohn)							
Gep-ER 10-50	n	28	27	28	28	28	28
	Mean	22.5	-1.9	-6.1	-8.4	-9.4	-9.0
	SE	0.55	0.6	0.8	1.0	1.1	1.1
Gep-ER 20-100	n	26	25	26	26	26	26
	Mean	21.8	-2.5	-5.0	-5.0	-5.7	-6.3
	SE	0.39	0.6	0.8	1.1	1.1	1.2
Gep-ER Total	n	54	52	54	54	54	54
	Mean	22.1	-2.2	-5.6	-6.8	-7.6	-7.7
	SE	0.35	0.4	0.6	0.8	0.8	0.8
Placebo	n	29	28	29	29	29	29
	Mean	23.1	-3.2	-5.2	-7.0	-7.8	-7.5
	SE	0.45	0.6	0.8	1.0	1.1	1.1
Gep 10-50 vs Pbo	p-value		0.109	0.469	0.336	0.279	0.333
Gep 20-100 vs Pbo	p-value		0.420	0.862	0.178	0.188	0.457
Gep Total vs. Pbo	p-value		0.157	0.738	0.853	0.923	0.875
Site 002 (Ferguson)							
Gep-ER 10-50	n	20	19	20	20	20	20
	Mean	23.0	-2.7	-3.7	-4.3	-5.3	-6.0
	SE	0.63	0.8	1.1	1.2	1.3	1.4
Gep-ER 20-100	n	14	12	14	14	14	14
	Mean	22.2	-3.1	-4.6	-6.9	-9.0	-8.8
	SE	0.35	1.0	1.3	1.4	1.5	1.7
Gep-ER Total	n	22	31	34	34	34	34
	Mean	22.7	-2.9	-4.1	-5.4	-6.8	-7.1
	SE	0.39	0.6	0.9	0.9	1.0	1.1
Placebo	n	18	18	18	18	18	18
	Mean	22.7	-2.9	-4.9	-5.9	-5.3	-5.6
	SE	0.57	0.8	1.2	1.2	1.3	1.5
Gep 10-50 vs Pbo	p-value		0.849	0.450	0.347	0.990	0.868
Gep 20-100 vs Pbo	p-value		0.910	0.867	0.660	0.073	0.158
Gep Total vs. Pbo	p-value		0.940	0.559	0.725	0.369	0.413
Combined Sites							
Gep-ER 10-50	n	48	46	48	48	48	48
	Mean	22.7	-2.3	-4.9	-6.4	-7.3	-7.5
	SE	0.42	0.5	0.7	0.8	0.8	0.9
Gep-ER 20-100	n	40	37	40	40	40	40
	Mean	21.9	-2.8	-4.8	-5.9	-7.4	-7.5
	SE	0.28	0.5	0.8	0.9	0.9	1.0
Gep-ER Total	n	88	83	88	88	88	88
	Mean	22.3	-2.5	-4.8	-6.1	-7.2	-7.4
	SE	0.27	0.3	0.5	0.6	0.6	0.7
Placebo	n	47	46	47	47	47	47
	Mean	22.9	-3.1	-5.1	-6.5	-6.5	-6.5
	SE	0.35	0.5	0.7	0.8	0.9	0.9
Gep 10-50 vs Pbo	p-value		0.256	0.830	0.931	0.488	0.460
Gep 20-100 vs Pbo	p-value		0.717	0.807	0.667	0.506	0.473
Gep Total vs. Pbo	p-value		0.362	0.759	0.702	0.514	0.451

[Source: CSR CN105-078 Table 23 and Appendix F 7.1.1-1, 7.1.1-1A, 7.1.1-3, 7.1.1-3A, 7.1.1-4, 7.1.1-4A, 7.1.1-6, and 7.1.1-6A]
P-values and LS Means from ANOVA model with factors for treatment and center, including the interaction term (combined sites), or one-way ANOVA model with a factor for treatment (single site analysis).

Table 30 presents the change from baseline in the CGI responders by visit and treatment group for the ITT population (LOCF).

For combined sites, there were no statistically significant differences between gepirone ER and placebo at any time point for CGI Responder rate in the ITT population (LOCF analysis).

The by-center analysis showed significant differences at the Ferguson site for the gepirone ER 20-100 dose group at Week 4 (57% vs. 17%, p=0.019) and at Week 6/Endpoint (64% vs. 28%, p=0.042). Also, the low dose group differed significantly from placebo at Week 3 at the Cohn site (61% vs. 34%, p=0.049).
In the OC/ITT analysis, the percentage of CGI responders was significantly greater for high dose gepirone ER compared to placebo at Week 4. Similar positive trends and differences were noted in other analyses, including the LOCF analysis for the Evaluable population.

Table 30: CGI Responders at Each Visit (ITT/LOCF) — Study CN105-078

Treatment		Visit				
		Week 1	Week 2	Week 3	Week 4	Week 6/ET
Site 0001 (Cohn)						
Gep-ER 10-50	N	27	28	28	28	28
	n	0	8	17	18	16
	%	0%	29%	61%	64%	57%
Gep-ER 20-100	N	25	26	26	26	26
	n	3	7	7	9	13
	%	12%	27%	27%	35%	50%
Gep-ER Total	N	52	54	54	54	54
	n	3	15	24	27	29
	%	6%	28%	44%	50%	54%
Placebo	N	28	29	29	29	29
	n	1	5	10	14	13
	%	4%	17%	34%	48%	45%
Gep 10-50 vs Pbo	p-value	0.326	0.312	0.049	0.227	0.357
Gep 20-100 vs Pbo	p-value	0.251	0.390	0.548	0.310	0.704
Gep Total vs. Pbo	p-value	0.669	0.287	0.382	0.882	0.443
Site 0002 (Ferguson)						
Gep-ER 10-50	N	19	20	20	20	20
	n	0	3	4	7	8
	%	0%	15%	20%	35%	40%
Gep-ER 20-100	N	12	14	14	14	14
	n	1	4	6	8	9
	%	8%	29%	43%	57%	64%
Gep-ER Total	N	31	34	34	34	34
	n	1	7	10	15	17
	%	3%	21%	29%	44%	50%
Placebo	N	18	18	18	18	18
	n	0	4	5	3	5
	%	0%	22%	28%	17%	28%
Gep 10-50 vs Pbo	p-value	--	0.572	0.578	0.206	0.434
Gep 20-100 vs Pbo	p-value	0.221	0.685	0.381	0.019	0.042
Gep Total vs. Pbo	p-value	0.446	0.892	0.902	0.050	0.126
Combined Sites						
Gep-ER 10-50	N	46	48	48	48	48
	n	0	11	21	25	24
	%	0%	23%	44%	52%	50%
Gep-ER 20-100	N	37	40	40	40	40
	n	4	11	13	17	22
	%	11%	28%	33%	43%	55%
Gep-ER Total	N	83	88	88	88	88
	n	4	22	34	42	46
	%	5%	25%	39%	48%	52%
Placebo	N	46	47	47	47	47
	n	1	9	15	17	18
	%	2%	19%	32%	36%	38%
Gep 10-50 vs Pbo	p-value	0.326	0.639	0.193	0.086	0.227
Gep 20-100 vs Pbo	p-value	0.115	0.354	0.950	0.576	0.128
Gep Total vs. Pbo	p-value	0.467	0.442	0.435	0.196	0.123

P-values from CMH test; CGI Responder = "much improved" or "very much improved."
[*Source: CSRS CN105-078 Table 24 and Appendices F 7.1.2-1, 7.1.2-1A, 7.1.2-2, and 7.1.2-2A*]

Table 31 presents the results of the secondary efficacy endpoints for the ITT population (LOCF). Overall, the results favored high dose gepirone ER over placebo for all secondary efficacy variables, with significant differences detected for CGI global improvement (p=0.041) and HAMD Factor V (p=0.017).

Table 31: Summary of Secondary Efficacy Results at Endpoint (ITT/LOCF) — Study CN105-078

Efficacy Variable	End of Treatment Outcome			Pairwise Tests (p-values)	
	Gep Low (N = 48)	Gep High (N = 40)	Placebo (N = 47)	Low vs. P	High vs. P
HAMD-25 CFB	-9.8	-10.1	-8.7	0.492	0.406
HAMD-28 CFB	-11.2	-11.8	-10.2	0.546	0.403
HAMD-Item1 CFB	-1.0	-1.1	-0.7	0.169	**0.082**
CGI (severity) CFB	-0.8	-1.0	-0.7	0.630	0.186
CGI (global improvement)	2.8	2.5	3.0	0.386	**0.041**
% Responders (HAMD-17)	33%	35%	28%	0.533	0.465
MADRS CFB	-8.7	-9.9	-7.1	0.457	0.239
HAMD Factor 1 CFB (anxiety/somatization)	-2.2	-2.0	-1.9	0.513	0.733
HAMD Factor V CFB	-2.9	-3.5	-2.1	0.122	**0.017**
HAMD Factor VI CFB	-0.9	-0.7	-1.3	0.282	0.130

[*Source: CSR: CN105-078 Final Report Appendix F Tables 7.1.2-10, 7.1.2-27, 7.2.1-27, 7.2.1-3, 7.2.2-3, 7.2.2-27, 7.2.2-75, 7.2.2-99, 7.2.3-1 and 7.2.4-3*]
LS means and p-values from ANOVA (with treatment, center, and treatment-by-center interaction terms); CMH test for % responders
HAMD-17 Responder = 50% improvement from baseline.

As specified in the protocol, efficacy results were assessed for the presence of a treatment-by-center interaction (defined as $p \leq 0.10$). While no quantitative interactions were detected, qualitative interactions appeared to be present. Site 0001 showed trends favorable for the lower dose of gepirone ER and Site 0002 showed trends favorable for the higher dose of gepirone ER. Additionally, Site 0001 showed substantial placebo response mean HAMD-17 improvement. These differences may be due to a difference in clinical profiles of patients at the two study centers. When compared to site 0001, the subject population at Site 0002 consisted of a larger percent experiencing a recurrent episode of MDD, with depression manifesting with melancholia, with a longer mean duration of previous depressive episodes, and with a shorter mean duration of the current depressive episode.

The Sponsor performed post-hoc mixed model analyses for HAMD-17 and several other parameters, including the modified HAMD-17 (mHAMD-17), the HAMD core depression factor (Bech-6), HAMD Item 1, and the MADRS total score. Mixed-models analyses were performed under 2 different covariance structures: first-order autoregressive structure and compound symmetry for the ITT dataset. Based on either model, the high dose group showed significantly greater improvement than the placebo group in mHAMD-17 ($p \leq 0.012$), Bech-6 ($p \leq 0.032$), and HAMD-Item 1 ($p \leq 0.039$) scores. [Note: The mHAMD-17 scale replaces items related to insomnia, nausea and agitation with ratings of the opposite (or reverse) neuro-vegetative symptoms from the HAMD-25. This scale was an ad hoc modification developed by Organon, the NDA sponsor at the time the mixed model analysis was done.]

Conclusions

This study was a 2-center, randomized, double-blind, placebo-controlled, flexible dose trial evaluating gepirone ER at 2 dose levels (10-50 mg/day [low dose] and 20-100 mg/day [high dose]) in subjects with MDD during a 6-week treatment period and consisted of the following flaws.

The planned sample size (180 subjects) was determined to give at least 72% power to detect a difference in HAMD-17 total score of 3 points between gepirone ER and placebo, assuming a variance of 55. However, due to the study's premature termination, only 144 subjects (80%) were treated in this study, severely reducing the power to detect statistically significant differences to approximately 62%.

In addition, placebo response at the two study sites differed markedly. At the Ferguson site, the placebo response (based on CGI) was considerably lower than the high-dose gepirone ER response (28% vs. 65%), whereas the Cohn site showed response rates of 45% on placebo vs. 50% on high dose gepirone ER.

The study also employed inadequate doses of gepirone ER. The mean maximal modal doses were: 37.3 mg in the gepirone ER 10-50 mg group and 67.5 mg in the gepirone ER 20-100 mg group. By protocol, the primary efficacy analysis was a comparison of the pooled gepirone ER dose groups (10-50 mg/day and 20-100 mg/day) with placebo. This weakened the potential of the study to demonstrate antidepressant efficacy.

Overall, study CN105-078 is uninterpretable regarding the efficacy of gepirone ER and FDA concurs.

9.1.2CN105-083

Study Description

This was a 2-center, randomized, double-blind, placebo-controlled, flexible dose trial evaluating 2 doses of gepirone ER (10-50 mg/day [low dose] and 20-100 mg/day [high dose]) in subjects with MDD during a 6-week treatment period. The average dose in the Low Dose gepirone ER Group (10-50 mg/day) was 30.4 ± 7.0 mg/day and the mean maximal dose was 37.2 ± 11.1 mg/day, which is below the minimum effective dose of 40 mg/day. The average dose in the High Dose Group (20-100 mg/day) was 57.1 ± 17.6 mg/day; the mean maximal dose was 70.3 ± 25.7 mg/day

Study Design

Randomized, double-blind, placebo-controlled, flexible dose, clinical study evaluating gepirone ER in the treatment of MDD. The study was conducted at 2 sites in the US and was designed to enroll 180 subjects to provide at leat 72% power. For business reasons, the study was terminated early by BMS short of the planned 180 subjects. At the time of termination, neither site had completed enrollment. Only 121 subjects were randomized (40 gepirone ER low-dose, 40

gepirone ER high-dose, and 41 placebo) and 117 subjects were treated, reducing the power of the study to approximately 53%.

After a 4-day to 4-week baseline period, eligible subjects were randomized to receive either placebo, gepirone ER 10-50 mg/day, or gepirone ER 20-100 mg/day for 6 weeks. A flexible dosing schedule was permitted by the protocol. The prescribed dosage was one tablet/day for the first two study days, increased to two tablets/day for Days 3 to 7. At Week 1, the dose was increased by one tablet every other day to four tablets/day for Weeks 2-3 unless adverse events supervened. The target dose for all dose groups was at least three tablets/day. After Week 3, dose titration was permitted up to five tablets/day depending on therapeutic response and tolerability. Subjects were evaluated at baseline (Day 0) and Weeks 1, 2, 3, 4, and 6 or early termination.

Entrance Criteria

Inclusion Criteria

Key inclusion criteria were:

- Patients at least 18 years of age
- Patients with a DSM-III-R diagnosis of MDD
- Patients with a baseline HAMD-17 total score ≥ 20

Efficacy Assessments

The co-primary efficacy parameters were the mean change from baseline in the HAMD-17 total score and the number (%) of CGI responders ("much improved" or "very much improved") at Week 6/ET based on the ITT/LOCF analysis.

Secondary Efficacy Measures

Key secondary efficacy measures, were as follows:

- Change from baseline on:

  - HAMD-25
  - HAMD-28
  - HAMD-Item 1
  - HAMD-17 responder rate
  - HAMD Factor I (anxiety/somatization)
  - HAMD Factor V (retardation)
  - HAMD Factor VI (sleep disturbance)
  - MADRS
  - CGI-severity score
  - CGI-global improvement

Statistical Methods

By protocol, the primary efficacy analysis was a comparison of the pooled gepirone ER dose groups (10-50 mg/day and 20-100 mg/day) with placebo. A secondary analysis compared each individual gepirone ER dose group with placebo.

The original intent of the protocol was to pool the two sites, if appropriate, with justification provided (e.g., a test of treatment-by-site interaction). For continuous data involving both centers, an ANOVA model was used with treatment and center as factors, and including the treatment-by-center interaction term. For by-center analysis, one-way ANOVA was used with treatment as a factor. CMH tests were used for categorical data, controlling for center. In addition, by-center analyses were presented.

Results

Overall, 117 subjects were randomized and treated (37 low dose, 39 high dose, 41 placebo).

As shown in the table below (Table 32), a total of 10 subjects (8.5%) were discontinued prematurely when the study was abruptly stopped and 43 subjects (36.8%) did not complete the 6-week treatment period. Reasons for drop-out included adverse events (13.5% low dose, 12.8% high dose, and 12.2% placebo) and a variety of other reasons such as lost to follow-up and withdrew consent. Most of the drop-outs (34 of 43 or 79%) occurred within the first 3 weeks of treatment.

Table 32: Subjects Discontinued by Reason (Combined Sites) — Study CN105-083

Number of Subjects		Treatment			
		Gepirone ER 10-50 mg/day	**Gepirone ER 20-100 mg/day**	**Placebo**	**Total**
Randomized		40	40	41	121
Treated		37	39	41	117
Discontinued	Total	14 (37.8%)	15 (38.5%)	14 (34.1%)	43 (36.8%)
	Adverse events	5 (13.5%)	5 (12.8%)	5 (12.2%)	15 (12.8%)
	Lack of efficacy	1 (2.7%)	2 (5.1%)	2 (4.9%)	5 (4.3%)
	Discontinued by BMS	2 (5.4%)	3 (7.7%)	5 (12.2%)	10 (8.5%)
	Other reason	6 (16.2%)	5 (12.8%)	2 (4.9%)	13 (11.1%)
Completed Treatment		23 (62.2%)	24 (61.5%)	27 (65.9%)	74 (63.2%)
[*Source: CSR CN105-083 Tables 10, 11 and 15, Appendix F Tables 6.1-1 and 8.1.1-1*]					

Table 33 presents the change from baseline in HAMD-17 score by visit and treatment group for the ITT population (LOCF). For combined sites, there were no statistically significant differences between the gepirone ER dose groups and the placebo group at any time point for the mean change from baseline to endpoint for the HAMD-17 total score in the ITT population (LOCF analysis). However, due to evidence of treatment by center interaction in the analysis of combined doses (p=0.073 at Week 4 and p=0.098 at Week 6/Endpoint), results of each center must be examined separately.

At Site 0002 (Cohn), greater improvement was noted in each gepirone ER dose group compared to the placebo group at all visits; the difference between the pooled doses vs. placebo approached statistical significance at Week 4 (p=0.089). Trends were reversed at Site 0001 (Fieve), with the placebo group showing marked improvement in HAMD-17 (43%) at the end of treatment.

Table 33: HAMD-17 Total Score: Change from Baseline at Each Visit (ITT/LOCF) — Study CN105-083

Treatment		Baseline	LS Mean Change from Baseline				
			Week 1	Week 2	Week 3	Week 4	Week 6/ET
Site 0001 (Fieve)							
Gep-ER 10-50	**n**	19	19	19	19	19	19
	Mean	24.8	-3.6	-6.5	-7.4	-9.3	-9.4
	SE	1.03	1.1	1.5	1.6	1.6	1.9
Gep-ER 20-100	**n**	21	21	21	21	21	21
	Mean	23.4	-3.4	-6.5	-7.3	-7.2	-8.0
	SE	0.41	1.1	1.4	1.5	1.5	1.8
Gep-ER Total	**n**	40	40	40	40	40	40
	Mean	24.1	-3.5	-6.5	-7.3	-8.2	-8.6
	SE	0.54	0.8	1.0	1.1	1.1	1.3
Placebo	**n**	23	23	23	23	23	23
	Mean	24.9	-4.7	-8.2	-8.7	-9.7	-10.7
	SE	0.83	1.0	1.4	1.4	1.5	1.7
Gep 10-50 vs Pbo	p-value		0.460	0.395	0.526	0.848	0.602
Gep 20-100 vs Pbo	p-value		0.390	0.396	0.489	0.246	0.270
Gep Total vs. Pbo	p-value		0.346	0.317	0.434	0.416	0.334
Site 0002 (Cohn)							
Gep-ER 10-50	**n**	17	17	17	17	17	17
	Mean	24.6	-4.5	-8.5	-9.8	-11.2	-10.2
	SE	0.95	1.1	1.4	1.5	1.6	1.7
Gep-ER 20-100	**n**	16	15	16	16	16	16
	Mean	22.7	-5.9	-8.7	-9.4	-10.8	-10.4
	SE	0.55	1.1	1.4	1.6	1.6	1.8
Gep-ER Total	**n**	33	32	33	33	33	33
	Mean	23.7	-5.1	-8.6	-9.6	-11.0	-10.3
	SE	0.57	0.8	1.0	1.1	1.1	1.2
Placebo	**n**	16	15	16	16	16	16
	Mean	22.7	-3.9	-6.4	-9.2	-7.6	-7.2
	SE	0.55	1.1	1.4	1.6	1.6	1.8
Gep 10-50 vs Pbo	p-value		0.732	0.302	0.816	0.114	0.237
Gep 20-100 vs Pbo	p-value		0.236	0.261	0.934	0.176	0.214
Gep Total vs. Pbo	p-value		0.391	0.210	0.853	**0.089**	0.159
Combined Sites							
Gep-ER 10-50	**n**	36	36	36	36	36	36
	Mean	24.8	-4.0	-7.5	-8.6	-10.3	-9.8
	SE	0.81	0.8	1.0	1.1	1.1	1.3
Gep-ER 20-100	**n**	37	36	37	37	37	37
	Mean	23.1	-4.6	-7.6	-8.4	-9.0	-9.2
	SE	0.33	0.8	1.0	1.1	1.1	1.3
Gep-ER Total	**n**	73	72	73	73	73	73
	Mean	23.9	-4.3	-7.5	-8.5	-9.6	-9.4
	SE	0.40	0.5	0.7	0.8	0.8	0.9
Placebo	**n**	39	38	39	39	39	39
	Mean	24.0	-4.3	-7.3	-9.0	-8.7	-8.9
	SE	0.56	0.8	1.0	1.1	1.1	1.3
Gep 10-50 vs Pbo	p-value		0.792	0.920	0.784	0.311	0.646
Gep 20-100 vs Pbo	p-value		0.764	0.846	0.684	0.830	0.902
Gep Total vs. Pbo	p-value		0.998	0.864	0.691	0.484	0.743

[*Source: CSR CN105-083 Table 23 and Appendix F Tables 7.1.1-1, 7.1.1-3, 7.1.1-1A, and 7.1.1-3A*]
P-values and LS Means from ANOVA model with factors for treatment and center, including the interaction term (combined sites), or one-way ANOVA model with a factor for treatment (single site analysis).

Table 34 presents the change from baseline in the CGI responders by visit and treatment group for the ITT population (LOCF). For combined sites, there were no statistically significant differences between gepirone ER and placebo at any time point for CGI Responder rate in the ITT population (LOCF analysis).

The by-center analysis showed trends in response rates favoring gepirone ER over placebo at Site 0002 (Cohn); significant differences were noted at Week 4 for the 20-100 gepirone ER group (69% vs. 25%, p=0.015), the 10-50 mg gepirone ER group (65% vs. 25%, p=0.024) and

the pooled dose groups (67% vs. 25%, p=0.007). The Cohn site showed similar positive findings for the high dose gepirone ER group at Week 4 in the OC/ITT analysis and the LOCF analysis of the Evaluable population. In the OC/ITT analysis, the percentage of CGI responders was significantly greater for high dose gepirone ER compared to placebo at Week 4

Table 34: CGI Responders at Each Visit (ITT/LOCF) — Study CN105-083

Treatment		Visit				
		Week 1	Week 2	Week 3	Week 4	Week 6/ET
Site 0001 (Fieve)						
Gep-ER 10-50	N	19	19	19	19	19
	n	0	4	6	7	8
	%	0%	21%	32%	37%	42%
Gep-ER 20-100	N	21	21	21	21	21
	n	2	5	6	6	9
	%	10%	24%	29%	29%	43%
Gep-ER Total	N	40	40	40	40	40
	n	2	9	12	13	17
	%	5%	23%	30%	33%	43%
Placebo	N	23	23	23	23	23
	n	4	7	9	9	11
	%	17%	30%	39%	39%	48%
Gep 10-50 vs Pbo	p-value	0.059	0.496	0.615	0.881	0.714
Gep 20-100 vs Pbo	p-value	0.453	0.626	0.466	0.466	0.744
Gep Total vs. Pbo	p-value	0.110	0.490	0.463	0.598	0.684
Site 0002 (Cohn)						
Gep-ER 10-50	N	17	17	17	17	17
	n	2	7	9	11	10
	%	12%	41%	53%	65%	59%
Gep-ER 20-100	N	15	16	16	16	16
	n	3	7	8	11	11
	%	20%	44%	50%	69%	69%
Gep-ER Total	N	32	33	33	33	33
	n	5	14	17	22	21
	%	16%	42%	52%	67%	64%
Placebo	N	15	16	16	16	16
	n	2	4	6	4	6
	%	13%	25%	38%	25%	38%
Gep 10-50 vs Pbo	p-value	0.895	0.332	0.381	0.024	0.228
Gep 20-100 vs Pbo	p-value	0.630	0.272	0.483	0.015	0.081
Gep Total vs. Pbo	p-value	0.839	0.240	0.362	0.007	0.088
Combined Sites						
Gep-ER 10-50	N	36	36	36	36	36
	n	2	11	15	18	18
	%	6%	31%	42%	50%	50%
Gep-ER 20-100	N	36	37	37	37	37
	n	5	12	14	17	20
	%	14%	32%	38%	46%	54%
Gep-ER Total	N	72	73	73	73	73
	n	7	23	29	35	38
	%	10%	32%	40%	48%	52%
Placebo	N	38	39	39	39	39
	n	6	11	15	13	17
	%	16%	28%	38%	33%	44%
Gep 10-50 vs Pbo	p-value	0.150	0.860	0.822	0.160	0.597
Gep 20-100 vs Pbo	p-value	0.811	0.706	0.937	0.278	0.378
Gep Total vs. Pbo	p-value	0.327	0.759	0.945	0.156	0.422

[*Source: CSR CN105-083 Table 24 and Appendices F 7.1.2-1, 7.1.2-1A, 7.1.2-2, and 7.1.2-2A*]
P-values from CMH test; CGI Responder = Much or Very Much Improved.

Table 35 presents the results of the secondary efficacy endpoints for the ITT population (LOCF). Overall, endpoint data from the combined study sites showed no significant difference between either gepirone ER dose and placebo for any of the secondary efficacy variables. In view of the qualitative interactions noted for the primary efficacy variables, results for Site 0002 (Cohn) were presented separately and showed positive results for the high dose gepirone ER group for several secondary efficacy variables in the Evaluable population]

Table 35: Summary of Secondary Efficacy Results at Endpoint (ITT/LOCF) — Study CN105-083

Efficacy Variable	End of Treatment Outcome			Pairwise Tests (p-values)	
	Gep Low (N = 36)	Gep High (N = 37)	Placebo (N = 39)	Low vs. P	High vs. P
HAMD-25 CFB	-11.9	-12.1	-11.3	0.813	0.739
HAMD-28 CFB	-14.1	-14.0	-13.2	0.763	0.777
HAMD-Item1 CFB	-1.1	-1.1	-1.1	0.833	0.773
CGI (severity) CFB	-1.2	-1.2	-1.2	0.936	0.983
CGI (global improvement)	2.8	2.6	2.8	0.961	0.508
% Responders (HAMD-17)	47%	43%	44%	0.737	0.979
MADRS CFB	-11.5	-13.4	-12.3	0.800	0.723
HAMD Factor 1 CFB (anxiety/somatization)	-2.6	-2.2	-2.4	0.733	0.695
HAMD Factor V CFB	-3.4	-3.6	-3.1	0.680	0.534
HAMD Factor VI CFB	-1.7	-1.2	-1.7	0.965	0.343

[*Source: CSR CN105-083 Final Report Appendix F Tables 7.1.2-10, 7.1.2-27, 7.2.1-27, 7.2.1-3, 7.2.2-3, 7.2.2-27, 7.2.2-75, 7.2.2-99, 7.2.3-1 and 7.2.4-3*]
LS means and p-values from ANOVA (with treatment, center, and treatment-by-center interaction terms); CMH test for % responders
HAMD-17 Responder = 50% improvement from baseline.

As specified in the protocol, results for the primary endpoints were assessed for the presence of a treatment-by-center interaction (defined as $p \leq 0.10$). For the primary analysis of HAMD-17 (ITT/LOCF) at Week6/Endpoint, a significant treatment-by-center interaction was noted in the model comparing the pooled gepirone ER dose groups to placebo (p=0.098). This interaction was found to be qualitative, with treatment effects in opposite directions at the two sites.

Placebo response was unusually high at Site 0001 (a reduction of 10.7 in HAMD-17 at endpoint, or 43%) compared to Site 0002 (reduction of 7.2, or 32%). Site 0001 also included more subjects in the ITT dataset (n=63) than in site 0002 (n=47). Consequently, it is reasonable to consider the sites independently. For Evaluable subjects at study Site 0002, the high dose gepirone ER group showed significant effects for CGI responders and HAMD-17 change from baseline based on the OC analyses.

Conclusions

This was a 2-center, randomized, double-blind, placebo-controlled, flexible dose trial evaluating 2 doses of gepirone ER (10-50 mg/day [low dose] and 20-100 mg/day [high dose]) in subjects with MDD during a 6-week treatment period. FDA and FKB agree that CN105-083 is a failed trial for the following reasons:

The study was terminated early with a small sample size, resulting in very low power. The number of subjects planned for this trial (60 per treatment group) would have afforded 72% power to detect a 3-point difference in the HAMD-17 total score at the 5% significance level. As a result of the BMS decision to stop development of gepirone ER, the study was terminated after only 117 subjects (65%) were treated, which reduced the power to detect statistically significant differences to only 53%.

Additionally, the LOCF method of analysis carries forward data for drop-outs to subsequent visits, thereby diluting the treatment effect. The early termination of the study caused disruptions in treatment and a relatively high discontinuation rate in the first 2 weeks of the trial, approximately 20%.

Overall, the study is uninterpretable and inadequate as a basis for efficacy conclusions because of its relatively small sample size, low range of gepirone ER doses employed, high drop-out rate in the first 3 weeks (partly due to early termination), and evidence of treatment-by-center interaction. Data from one of the two study sites (site 0002) suggest that high dose gepirone ER had activity in the OC dataset, but this finding for a secondary variable in a small subset is likely to be spurious. Overall, study CN105-083 is uninterpretable regarding the efficacy of gepirone ER.

9.1.3CN105-052

Study Description

This was a 2-center, 8-week, randomized, double-blind, active-controlled, flexible dose study evaluating gepirone ER (20-60 mg/day), fluoxetine (10-40 mg/day), and placebo in subjects with non-psychotic MDD. The maximum modal dose of gepirone ER was 43.4 ± 17.6 mg/day. The maximum modal dose of fluoxetine was 28.7±19.1 mg/day. In other studies, the Sponsor has shown that 40 mg/day is the minimum effective antidepressant dose of gepirone ER. Thus, at the time of study termination, subjects were at the lowest end of the therapeutic range for each of the study drugs.

Study Design

Randomized, double-blind, active-controlled, flexible dose, clinical study evaluating gepirone ER in the treatment of subjects with MDD. The study was conducted at 2 site in the US and was designed to enroll 240 subjects, as this would provide 80% power. However, BMS terminated the program after only 111 subjects had been randomized, decreasing the power to approximately 43%.

After a 4-day to 4-week baseline phase, eligible subjects were randomized by site to receive either gepirone ER tablets 20-60 mg (amended from 10-40 mg/day), fluoxetine 20 mg/day (amended from 20-80 mg/day), or placebo for 8 weeks (56 days). Study drugs were administered in a double-dummy, flexible dose manner. Subjects were evaluated at baseline (Day 0) and Weeks 1, 2, 3, 4, 6 and 8.

Entrance Criteria

Inclusion Criteria

Key inclusion criteria were:

- Patients at least 18 years of age;

- Patient with a diagnosis of MDD by DSM-III-R (single episode or recurrent)
- Patients with a baseline HAMD-17 total score $\geq$ 20

Efficacy Assessments

Primary Efficacy Measure

The co-primary efficacy parameters were the mean change from baseline in the HAMD-17 total score and the number (%) of CGI responders ("much improved" or "very much improved") at Week 8/End of Treatment based on the ITT/LOCF analysis.

Secondary Efficacy Measures

Key secondary efficacy measures, were as follows:

- Change from baseline on:

- HAMD-25
- HAMD-28
- HAMD-Item 1
- HAMD-17 responder rate
- MADRS
- CGI-severity score
- CGI-global improvement

Statistical Methods

The statistical plan for this study stipulated: "Data for the centers will be pooled, if appropriate. Justification for pooling will be provided, i.e., the statistical model will include center, treatment, and center-by treatment interaction effects." For continuous data, ANOVA was carried out with treatment and center as factors, and including the treatment-by-center interaction term. CMH tests were performed for categorical data. Comparisons between each active treatment group versus placebo were performed without adjusting for multiple comparisons.

Results

Overall, a total of 111 subjects had been randomized and 110 were treated (46% of the requisite sample size). The ITT population comprised 108 subjects with post-baseline data: 35 subjects gepirone ER, 36 fluoxetine, and 37 placebo.

As shown in Table 36, 47 subjects (42.3%) did not complete the 8-week treatment did not complete the 8-week treatment period. The discontinuation rate was higher in the placebo group (50.0%) than in the gepirone ER or fluoxetine groups (both 38.9%), primarily due to lack of efficacy. Adverse events were most common in the gepirone ER group (16.7%) compared to fluoxetine (5.6%) and placebo (2.6%).

Table 36: Subjects Discontinued by Reason (Combined Sites) — Study CN105-052

Number of Subjects		Treatment			
		Gepirone ER	Fluoxetine	Placebo	Total
Randomized		36	37	38	111
Treated		36	36	38	110
Discontinued	Total	14 (38.9%)	14 (38.9%)	19 (50.0%)	47 (42.7%)
	Adverse events	6 (16.7%)	2 (5.6%)	1 (2.6%)	9 (8.2%)
	Lack of efficacy	4 (11.1%)	4 (11.1%)	8 (21.1%)	16 (14.5%)
	Discontinued by BMS	2 (5.6%)	3 (8.3%)	6 (15.8%)	11 (10.0%)
	Other reason	2 (5.6%)	5 (13.9%)	4 (10.5%)	11 (10.0%)
Completed Treatment		22 (61.1%)	22 (61.1%)	19 (50.0%)	63 (57.3%)

[*Source: CSR CN105-052 Tables 5 and 8*]

Table *37* presents the change from baseline in the HAMD-17 score y visit and treatment group for the ITT population (LOCF). There were no statistically significant differences detected between either active drug and placebo, based on the change from baseline in HAMD-17 total scores.

Table 37: HAMD-17 Total Score: Change from Baseline at Each Visit (ITT/LOCF) — Study CN105-052

Treatment		Baseline	LS Mean Change from Baseline					
			Week 1	Week 2	Week 3	Week 4	Week 6	Week 8/ET
Gep-ER	**n**	35	34	35	35	35	35	35
	Mean	25.5	-2.6	-6.0	-8.4	-8.9	-10.0	-11.0
	SE	0.54	0.7	0.9	1.1	1.2	1.4	1.5
Fluoxetine	**n**	36	33	36	36	36	36	36
	Mean	25.2	-2.4	-4.4	-7.6	-9.2	-10.7	-11.0
	SE	0.43	0.7	0.9	1.0	1.2	1.3	1.5
Placebo	**n**	37	37	37	37	37	37	37
	Mean	25.2	-2.7	-5.6	-7.7	-9.2	-9.7	-10.5
	SE	0.54	0.7	0.9	1.0	1.2	1.3	1.4
Gepirone ER vs. Pbo	p-value		0.928	0.736	0.660	0.880	0.905	0.825
Fluoxetine vs. Pbo	p-value		0.744	0.322	0.934	0.996	0.600	0.798

LS Means and p-values for combined sites is based on ANOVA with effects for treatment, center and treatment by center interaction.
[Source: *CSR CN105-052 Table 16, Appendix F Tables 7.1.1-1 and 7.1.1-3*]

Table 38 presents the change from baseline in the CGI reponders by visit and treatment group for the ITT population (LOCF). No statistically significant differences were detected between either active drug and placebo based on the CGI responder rate.

Table 38: Number (%) CGI Responders at Each Visit (ITT/LOCF) — Study CN105-052

Treatment		Visit					
		Week 1	Week 2	Week 3	Week 4	Week 6	Week 8/ET
Gep-ER	**N**	34	35	35	35	35	35
	n	3	14	20	21	21	21
	%	9%	40%	57%	60%	60%	60%
Fluoxetine	**N**	33	36	36	36	36	36
	n	5	11	18	20	22	24
	%	15%	31%	50%	56%	61%	67%
Placebo	**N**	37	37	37	37	37	37
	n	6	9	20	20	22	21
	%	16%	24%	54%	54%	59%	57%
Gepirone ER vs. Pbo	p-value	0.356	0.152	0.792	0.614	0.965	0.783
Fluoxetine vs. Pbo	p-value	0.944	0.568	0.699	0.900	0.897	0.395

[Source: *CSR CN105-052 Table 17, Appendix F Tables 7.1.2-1*]; P-values from CMH test; CGI Responder = Much or Very Much Improved

Table 39 presents the results of the secondary efficacy endpoint for the ITT population (LOCF). No statistically significant differences were detected between either active treatment and placebo, based on any of the secondary efficacy variables.

Table 39: Summary of Secondary Efficacy Results at Endpoint — Study CN105-052

Efficacy Variable	End of Treatment Outcome			P-values	
	Gepirone N=35	Fluoxetine N=36	Placebo N=37	G vs P	F vs P
HAMD-28 CFB	-15.8	-17.6	-15.8	0.984	0.564
HAMD-25 CFB	-13.1	-14.0	-13.0	0.967	0.682
HAMD-Item1 CFB	-1.3	-1.5	-1.3	0.897	0.486
HAMD-17 Responders	49%	50%	43%	0.653	0.570
CGI (severity) CFB	NA	NA	NA	NA	NA
CGI (global improvement)	NA	NA	NA	NA	NA
MADRS CFB	-15.7	-15.5	-13.5	0.475	0.522

NA = not analyzed
[*Source: CSR CN105-052 Table 18, 19, 20, 21; Appendix F Tables 7.2.1-3, 7.2.2-3, 7.2.1-26, 7.2.3-1, 7.2.4-3*]

Conclusions

This study was a 2-center, 8-week, randomized, double-blind, active-controlled, flexible dose study evaluating gepirone ER (20-60 mg/day), fluoxetine (10-40 mg/day), and placebo in subjects with non-psychotic MDD. This study is considered uninterpretable due to the following factors.

First, the study was prematurely terminated by BMS with less than 40 subjects per treatment group, only 46% of the required sample size, thus reducing the power to detect treatment effects from 80% to approximately 43%.

Second, the placebo response rate was relatively high: 57% of placebo subjects were CGI responders at Week 8/Endpoint. This would tend to obscure evidence of response to active treatment.

Third, the dose of gepirone used in this study (average maximum dose of 43.4 mg/day) was relatively low within the proposed therapeutic dose range (40-80 mg/day).

Finally, the study was unable to differentiate the well-known treatment effects of the active control product (fluoxetine) given at therapeutic doses (10-40 mg) from placebo.

The Sponsor and FDA agree that study CN105-052 lacks assay sensitivity and is therefore an uninterpretable trial. The inability to differentiate between the effects of the active control product (fluoxetine) given at therapeutic doses (10-40 mg) and placebo makes this a failed study. Several factors contributed to the failure of the study:

- *The study was prematurely terminated by BMS with less than 40 subjects per treatment group, only 46% of the required sample size, thus reducing the power to detect treatment effects (under assumptions on p. 56 of the CSR) from 80% to 43%.*
- *The placebo response rate was extremely high: 57% of placebo subjects were CGI responders at Week 8/Endpoint. This tends to obscure evidence of response to active treatment.*
- *The dose of gepirone-ER used in this study (average maximum dose of 43.4 mg/day) was relatively low within the proposed therapeutic dose range (40-80 mg/day).*

9.1.4 CN105-053

Study Description

This was a 2-center, randomized, double-blind, active-controlled, flexible dose study evaluating gepirone ER (10-60 mg/day), imipramine (50-200 mg/day) and placebo over an 8-week treatment period in subjects with non-psychotic MDD. The original dose of gepirone ER (10–40 mg/day) was amended to 10-60 mg/day after 5 months. The average dose administered at both sites was 50.4±13.9 mg/day. However, while the average dose administered at the both sites and Site 0001 at 53.4 ± 11.5 mg/day were above the minimum effective dose of gepirone, 40 mg/day, the average dose administered at Site 0002 was only 42.0 ± 16.6 mg/day.

Study Design

Randomized, double-blind, active-controlled, flexible dose, clinical study evaluating gepirone ER in the treatment of subjects with MDD. The study was conducted at 2 sites in the US and was designed to enroll 240 subjects, as this would provide 80% power. BMS terminated the program early, after 170 subjects were randomized (58 gepirone, 56 imipramine, and 56 placebo), reducing the power to approximately 63%.

After a 4-day to 4-week baseline period, eligible subjects were randomized by site to receive either placebo, gepirone ER tablets (10-60 mg), or imipramine (50-200 mg) once per day in the morning for

8 weeks (56 days). Study drugs were administered in a double-dummy, flexible dose manner. Subjects were evaluated at baseline (Day 0) and Weeks 1, 2, 3, 4, 6 and 8 or early termination.

Entrance Criteria

Key inclusion criteria were:

- Patients at least 18 years of age;
- Patient with a diagnosis of MDD by DSM-III-R (single episode or recurrent)
- Patients with a baseline HAMD-17 total score $\geq$ 20

Efficacy Assessments

Primary Efficacy Measure

The co-primary efficacy parameters were the mean change from baseline in the HAMD-17 total score and the number (%) of CGI responders ("much improved" or "very much improved") at Week 8/End of Treatment based on the ITT/LOCF analysis.

Secondary Efficacy Measures

Key secondary efficacy measures, were as follows:

- Change from baseline on:

  - HAMD-25
  - HAMD-28
  - HAMD-Item 1
  - HAMD-17 responder rate
  - MADRS
  - CGI-severity score
  - CGI-global improvement

Statistical Methods

The statistical plan for this study stipulated: "Data for the centers will be pooled, if appropriate. Justification for pooling will be provided, i.e., the statistical model will include center, treatment, and center-by-treatment interaction effects." (See Protocol page 32). For analysis of continuous data from both centers, the ANOVA model included effects for treatment and center, as well as the interaction term. For all by-center analyses, a one-way ANOVA model was used with treatment as a factor. CMH tests were performed for categorical data.

Results

Overall, a total of 170 subjects had been randomized and 168 were treated (70% of the requisite sample size). The ITT population comprised 166 subjects with post-baseline data: 120 subjects at

the Feiger site (41 gepirone ER, 39 imipramine, and 40 placebo) and 46 subjects at the Gelenberg site (15 gepirone ER, 15 imipramine, and 16 placebo).

As shown in Table 40, a total of 78 (46.4%) of subjects did not complete the study. The pattern of drop-outs differed considerably between the 2 study sites. At the Feiger site (001), the discontinuation rate was higher in the placebo group (67.5%) than in the gepirone ER or imipramine groups (38.1% and 38.5%, respectively), primarily due to lack of efficacy; adverse events were most common in the imipramine group (28.2%) compared to gepirone ER (4.8%) and placebo (2.5%). By contrast, discontinuations at the Gelenberg site were most frequent in the gepirone ER group (50.0%), primarily due to adverse events; drop-outs due to adverse events were higher in gepirone ER and imipramine groups (31.3% and 26.7%, respectively) than in the placebo group (6.3%).

Notably, although Site 001 (Feiger) completed enrollment, due to the study's premature termination, only 39% of the planned enrollment was accomplished at site 002 (Gelenberg), limiting each treatment group to only 15-16 subjects. Additionally, there were major differences between study sites in demographic and baseline characteristics. Compared to the Gelenberg site, subjects in the Feiger site were younger (39 vs. 43 years), p=0.027; had fewer previous depressive episodes (60% vs. 81%), p=0.011; and had lower HAMD-17 baseline scores (23.7 vs. 25.1), p=0.006.

Table 40: Subjects Discontinued by Reason and Study Site — Study CN105-053

Number of Subjects		Treatment			
		Gepirone ER	Imipramine	Placebo	Total
Feiger (Site 001)					
Randomized		42	41	40	123
Treated		42	39	40	121
Discontinued	Total	16 (38.1%)	15 (38.5%)	27 (67.5%)	58 (47.9%)
	Adverse events	2 (4.8%)	11 (28.2%)	1 (2.5%)	14 (11.6%)
	Lack of efficacy	11 (26.2%)	1 (2.6%)	24 (60%)	36 (29.8%)
	Discont by BMS	0	0	0	0
	Other reason	3 (7.1%)	3 (7.7%)	2 (5.0%)	8 (6.6%)
Completed Treatment		26 (61.9%)	24 (61.5%)	13 (32.5%)	63 (52.1%)
Gelenberg (Site 002)					
Randomized		16	15	16	47
Treated		16	15	16	47
Discontinued	Total	8 (50.0%)	5 (33.3%)	7 (43.8%)	20 (42.6%)
	Adverse events	5 (31.3%)	4 (26.7%)	1 (6.3%)	10 (21.3%)
	Lack of efficacy	1 (6.3%)	0	4 (25.0%)	5 (10.6%)
	Discont by BMS	1 (6.3%)	0	1 (6.3%)	2 (4.3%)
	Other reason	1 (6.3%)	1 (6.7%)	1 (6.3%)	3 (6.4%)
Completed Treatment		8 (50.0%)	10 (66.7%)	9 (56.3%)	27 (57.4%)
Combined Sites					
Randomized		58	56	56	170
Treated		58	54	56	168
Discontinued	Total	24 (41.4%)	20 (37.0%)	34 (60.7%)	78 (46.4%)
	Adverse events	7 (12.1%)	15 (27.8%)	2 (3.6%)	24 (14.3%)
	Lack of efficacy	12 (20.7%)	1 (1.9%)	28 (50.0%)	41 (24.4%)
	Discont by BMS	1 (1.7%)	0	1 (1.8%)	2 (1.2%)
	Other reason	4 (7.1%)	4 (7.4%)	3 (5.4%)	11 (6.5%)
Completed Treatment		34 (58.6%)	34 (63.0%)	22 (39.3%)	90 (53.6%)

[*Source: CSR CN105-053 Table 5, 6 and 9*]

Table 41 presents the change from baseline in HAMD-17 score by visit, treatment group, and sites for the ITT population (LOCF).

In the Feiger study (site 001), reductions from baseline in HAMD-17 ratings were significantly greater in the gepirone ER group compared to placebo at Week 8/Endpoint ($p = 0.049$). The reductions in HAMD-17 were also significantly greater for imipramine than placebo at Week 4 ($p=0.024$), Week 6 ($p=0.014$) and Week 8/Endpoint ($p=0.017$).

The Gelenberg study (site 0002), with only 15-16 subjects per treatment group, showed no statistically significant differences between either of the active treatments and placebo.

For the combined sites, trends in HAMD-17 favored each active drug over placebo, but differences did not achieve statistically significance at the Week 8/Endpoint or at any other visits except Week 2 (imipramine vs. placebo, $p=0.041$).

The above results are consistent with the other analyses, including the OC analysis in the ITT population and the LOCF and OC analyses in the Evaluable population

Table 41: HAMD-17 Total Score: Change from Baseline at Each Visit (ITT/LOCF) − Study CN105-053

Treatment		Baseline	LS Mean Change from Baseline					
			Week 1	Week 2	Week 3	Week 4	Week 6	Week 8/ET
Feiger (Site 001)								
gepirone ER	**n**	41	40	41	41	41	41	41
	Mean	23.7	-4.7	-5.5	-7.2	-8.2	-10.8	-10.1
	SE	0.52	0.7	0.9	1.0	1.0	1.1	1.2
imipramine	**n**	39	39	39	39	39	39	39
	Mean	23.6	-4.5	-7.0	-9.3	-9.4	-11.8	-10.9
	SE	0.48	0.7	0.9	1.0	1.0	1.2	1.2
placebo	**n**	40	38	40	40	40	40	40
	Mean	23.9	-3.9	-4.9	-6.8	-6.0	-7.7	-6.8
	SE	0.46	0.7	0.9	1.0	1.0	1.2	1.2
gepirone ER vs pbo	p-value		0.455	0.649	0.746	0.129	**0.063**	**0.049**
imipramine vs. pbo	p-value		0.572	**0.096**	**0.092**	**0.024**	**0.014**	**0.017**
Gelenberg (Site 002)								
gepirone ER	**n**	15	15	15	15	15	15	15
	Mean	24.6	-3.1	-7.3	-9.0	-9.1	-10.3	-9.3
	SE	0.62	1.3	1.5	1.7	2.0	2.5	2.5
imipramine	**n**	15	15	15	15	15	15	15
	Mean	25.7	-3.1	-7.5	-7.9	-10.9	-10.9	-12.2
	SE	0.54	1.3	1.5	1.7	2.0	2.5	2.5
placebo	**n**	16	15	16	16	16	16	15
	Mean	25.0	-2.8	-4.7	-6.4	-9.1	-10.2	-11.2
	SE	0.75	1.3	1.1	1.7	2.0	2.4	2.5
gepirone ER vs. pbo	p-value		0.837	0.226	0.289	0.998	0.966	0.589
imipramine vs. pbo	p-value		0.837	0.182	0.534	0.527	0.843	0.776
Combined Sites								
gepirone ER	**n**	56	55	56	56	56	56	56
	Mean	23.9	-3.9	-6.4	-8.1	-8.7	-10.5	-9.7
	SE	0.41	0.7	0.8	1.0	1.0	1.2	1.2
imipramine	**n**	54	54	54	54	54	54	54
	Mean	24.2	-3.8	-7.3	-8.6	-10.2	-11.3	-11.5
	SE	0.39	0.7	0.9	1.0	1.1	1.2	1.2
placebo	**n**	56	53	56	56	56	56	56
	Mean	24.2	-3.3	-4.8	-6.6	-7.6	-8.9	-9.0
	SE	0.39	0.7	0.8	1.0	1.0	1.2	1.2
gepirone ER vs. pbo	p-value		0.568	0.195	0.274	0.448	0.343	0.687
imipramine vs. pbo	p-value		0.633	**0.041**	0.153	**0.080**	0.160	0.144

LS Means and p-values for combined sites is based on ANOVA with effects for treatment, center and treatment by center interaction.
[*Source: CSR CN105-053 Table 17, Appendix F Tables 7.1.1-3 and 7.1.1-6*]

Table 42 presents the change from baseline in the CGI responders by visit, treatment group, and site for the ITT population (LOCF).

In the Feiger study (site 001), the gepirone ER group had a 56% CGI response rate at Week 8/Endpoint, which was statistically significant compared to placebo ($p = 0.034$**)**. The imipramine group also showed significantly higher response rates than placebo at Week 8/Endpoint ($p=0.001$) as well as Week 3 ($p=0.034$), Week 4 ($p=0.003$) and Week 6 ($p=0.002$).

The Gelenberg site showed no significant differences in response rates between either of the active treatments and placebo at any study visit; by Week 8, the placebo response rate was unusually high (56%) compared to the rate observed at the Feiger site (33%).

For the combined sites, trends in response rates favored each active treatment over placebo; differences were significant for imipramine at Weeks 3 through Week8/Endpoint ($p < 0.001$).

Table 42: Number (%) CGI Responders at Each Visit (ITT/LOCF) — Study CN105-053

Treatment		Visit					
		Week 1	Week 2	Week 3	Week 4	Week 6	Week 8/ET
Feiger (Site 001)							
gepirone ER	**N**	40	41	41	41	41	41
	n	9	15	16	19	24	23
	%	23%	37%	39%	46%	59%	56%
imipramine	**N**	39	39	39	39	39	39
	n	7	18	24	25	30	28
	%	18%	46%	62%	64%	77%	72%
placebo	**N**	38	40	40	40	40	40
	n	4	10	15	12	17	13
	%	11%	25%	38%	43%	33%	33%
gepirone ER vs. pbo	p-value	0.159	0.262	0.888	0.133	0.151	**0.034**
imipramine vs. pbo	p-value	0.355	**0.051**	**0.034**	**0.003**	**0.002**	**0.001**
Gelenberg (Site 002)							
gepirone ER	**N**	15	15	15	15	15	15
	n	1	4	8	7	9	7
	%	7%	27%	53%	47%	60%	47%
imipramine	**N**	15	15	15	15	15	15
	n	1	5	8	12	10	11
	%	7%	33%	53%	80%	67%	73%
placebo	**N**	15	16	16	16	16	16
	n	1	2	2	8	8	9
	%	7%	13%	13%	50%	50%	56%
gepirone vs. pbo	p-value	1.0	0.326	**0.017**	0.855	0.582	0.600
imipramine vs. pbo	p-value	1.0	0.173	**0.017**	0.086	0.355	0.328
Combined Sites							
gepirone ER	**N**	55	56	56	56	56	56
	n	10	19	24	26	33	30
	%	18%	34%	43%	46%	59%	54%
imipramine	**N**	54	54	54	54	54	54
	n	8	23	32	37	40	39
	%	15%	43%	59%	69%	74%	72%
placebo	**N**	53	56	56	56	56	56
	n	5	12	17	20	25	22
	%	9%	21%	30%	36%	45%	39%
gepirone ER vs. pbo	p-value	0.192	0.147	0.182	0.240	0.131	0.125
imipramine vs. pbo	p-value	0.398	**0.018**	**0.002**	**0.001**	**<0.001**	**<0.001**

[*Source: CSR CN105-053 Table 18, Appendix F Tables 7.1.2-1, 7.1.2-2*]

Table 43 presents the results of the secondary efficacy endpoints for the ITT population (LOCF).

The inconsistent results and imbalance in sample sizes for the two study sites is noteworthy. An unusually high placebo response was observed at the Gelenberg site (-11.2 points on the HAMD-17 scale, or a 45% reduction from the baseline mean of 25.0), compared to the Feiger site (-6.8 points, or 28% reduction from the baseline mean of baseline).

At the Feiger site, both gepirone ER and imipramine showed positive and statistically significant treatment effects at Week 8/Endpoint for nearly all efficacy variables. In only one instance (change from baseline MADRS) was the effect of imipramine significant when the effect of gepirone ER was not.

The results were interpreted according to the protocol and presented efficacy results for each study site. It was not considered appropriate to pool data from the two study sites due to the disparity in enrollment and highly inconsistent treatment effects. The Feiger site showed significant positive results for gepirone ER and imipramine compared to placebo; no significant treatment effects were detected in the Gelenberg site. According to the ANOVA analysis planned in the protocol, the HAMD-17 scores for the combined sites show no statistically significant effects for either drug.

The FDA did a post hoc re-analysis of data from the combined sites only and found that the mean reduction in HAMD-17 at Week 8/ET was significantly greater for imipramine compared to placebo ($p = 0.038$), but not for gepirone ER ($p = 0.190$). It is not clear what model was used; Fabre-Kramer was not able to replicate the p-values and question how the p-value for imipramine vs. placebo could have changed from 0.144 to 0.038.

Table 43: Summary of Secondary Efficacy Results at Endpoint by Site and Overall — Study CN105-053

Efficacy Variable	End of Treatment Outcome			P-values	
	Gepirone ER	Imipramine	Placebo	G vs P	I vs P
Feiger	N=41	N=39	N=40		
HAMD-28 CFB	-17.3	-17.6	-10.4	**0.006**	**0.005**
HAMD-25 CFB	-14.4	-14.6	-8.9	**0.007**	**0.006**
HAMD-Item1 CFB	-1.2	-1.4	-0.6	**0.010**	**0.001**
HAMD-17 Responders	44%	49%	33%	0.294	0.145
CGI (severity) CFB	-1.5	-1.7	-0.9	**0.031**	**0.004**
CGI (global improvement)	2.4	2.1	3.1	**0.012**	**0.001**
MADRS CFB	-11.9	-14.3	-8.7	0.179	**0.020**
Gelenberg	N=15	N=15	N=16		
HAMD-28 CFB	-13.5	-17.3	-14.8	0.798	0.614
HAMD-25 CFB	-11.3	-14.3	-12.9	0.709	0.724
HAMD-Item1 CFB	-0.7	-1.1	-1.3	0.163	0.550
HAMD-17 Responders	40%	67%	50%	0.582	0.355
CGI (severity) CFB	-1.1	-1.8	-1.4	0.529	0.340
CGI (global improvement)	2.5	2.4	2.7	0.718	0.638
MADRS CFB	-12.7	-16.5	-15.9	0.502	0.900
Combined	N=56	N=54	N=56		
HAMD-28 CFB	-15.4	-17.5	-12.6	0.266	0.055
HAMD-25 CFB	-12.9	-14.5	-10.9	0.330	0.084
HAMD-Item1 CFB	-1.0	-1.2	-0.9	0.929	0.200
HAMD-17 Responders	43%	54%	38%	0.551	0.084
CGI (severity) CFB	-1.3	-1.8	-1.1	0.535	**0.021**
CGI (global improvement)	2.4	2.2	2.9	0.110	**0.031**
MADRS CFB	-12.3	-15.4	-12.3	0.987	0.197

[*Source: CN105-053 Final Report Appendices F7.1, 7.2, and 7.4*]
LS means and p-values from ANOVA, with treatment and center as factors; CMH test for % responders

Conclusions

This was a 2-center, randomized, double-blind, active-controlled, flexible dose study evaluating gepirone ER (10-60 mg/day), imipramine (50-200 mg/day) and placebo over an 8-week treatment period in subjects with non-psychotic MDD.

The two study sites show conflicting results due, at least in part, to the imbalance in sample sizes, differences in demographics, the low dose of study drug at Site 0002, and high placebo response rates at Site 0002.

When the study was terminated, Site 0001 has exceeded enrollment (n=123), but Site 002, only enrolled 47 (39%) of the 120 required.

When the two sites were pooled and then analyzed, no significant pairwise differences (gepirone ER vs. placebo or imipramine vs. placebo) were detected for HAMD-17; a significant difference was noted for CGI response rates favoring imipramine over placebo (p=0.001). However, when

the two sites were evaluated separately, Site 0002 (with only 15-16 subjects per treatment group) showed no positive effects of either drug on the primary or secondary efficacy endpoints. Site 0001, on the other hand, showed significant treatment effects for both gepirone ER and imipramine on the co-primary efficacy endpoints and most of the secondary efficacy parameters.

9.1.5134004

Study Description

This was a 10-center, 3-arm, randomized, double-blind, active-controlled, study evaluating gepirone ER in comparison to placebo and fluoxetine over an 8-week treatment period in subjects diagnosed with MDD having "atypical features." The mean final dose of gepirone was 67.1 mg/day (±19.2), with 79.3% of subjects titrated to 60-80 mg/day. The mean final dose of fluoxetine was 34.1 mg/day (±9.2), with 70.3% of subjects at the top dose.

Study Design

Randomized, double-blind, active-controlled, clinical study evaluating gepirone ER in the treatment of MDD with "atypical features." The study was conducted at 10 sites in the U.S. 410 subjects were randomized and 409 received study medication (135 gepirone ER, 138 fluoxetine, and 136 placebo). A total of 391 subjects with post-baseline data comprised the ITT population (125 gepirone ER, 136 fluoxetine, and 130 placebo). The dose range for gepirone ER was 20-80 mg once daily, administered orally in the morning. After Day 7, the minimum dose was 40 mg/day. The dose range for fluoxetine was 20-40 mg once daily, administered orally in the morning. Blinding was accomplished by providing placebo in tablets or capsules identical in appearance to gepirone ER or fluoxetine, respectively. Subjects were evaluated at baseline (Day 0) and Weeks 1, 2, 3, 4, 6 and 8.

Entrance Criteria

Inclusion Criteria

Key inclusion criteria were:

- Patients 18-65 years of age
- Patients who met criteria MDD (DSM-IV code 296) with either a single episode (296.2x) or recurrent episodes (296.3x)
- Patients who suffered from moderate depression (296.x2) to severe depression without psychotic features
- Patients who had no more than a 20% decrease on the HAMD-25 total score between screening and baseline
- Patients who met criteria for DSM-IV MDD with Atypical Features Specifier as assessed using the Atypical Depression Diagnostic Scale (ADDS)

- Patients who had dysphoria for most days over the past 4 weeks
- Patients who had a current episode of MDD with atypical features lasting at least 3 months, whether the subject was diagnosed with one single episode or recurrent episodes

Efficacy Assessments

Primary Efficacy Measure

The primary efficacy endpoint was the change from baseline (CFB) in the HAMD-25 at week 8 or end of treatment (EOT) using LOCF approach in the ITT population. The HAMD-25 is the total score for items 1-18 and items 22-28 of the HAMD 31-item scale.

Secondary Efficacy Measures

Secondary efficacy measures, were as follows:

- Comparisons of gepirone ER vs. placebo: change from baseline on
  - HAMD-17
  - HAMD-28
  - HAMD-Item 1 (depressed mood)
  - CGI severity
  - CGI improvement
  - HAMD-25 responders
  - CGI responders
  - HAMD-25 remitters.
- Comparisons of gepirone ER vs. placebo and gepirone ER vs. fluoxetine: change from baseline on
  - HAMD hypersomnia/hyperphagia factor (sum of items 22-26 of HAMD-31)
  - HAMA total score.

Of these secondary efficacy variables, none was defined or designated as “key” in the protocol.

Statistical Methods

For the primary endpoint, an ANOVA model was applied, with treatment (gepirone ER vs. placebo) and site as factors. If the effect of gepirone ER was not significant, fluoxetine was compared to placebo (using the same ANOVA model for 2 treatment groups) in order to validate the outcome. As stated in the protocol: “The comparison of fluoxetine to placebo, a secondary objective, is made in order to validate the outcome of the comparison of gepirone ER and placebo and will be done only in the case gepirone ER fails to show superiority over placebo.”

While not a traditional approach, this was the pre-planned analysis. The intent of the protocol was that if fluoxetine was also not statistically different from placebo, the study was uninterpretable and further statistical evaluation was not warranted.

Secondary efficacy variables were analyzed as follows:

- *Comparisons of gepirone-ER vs. placebo* — change from baseline for HAMD-17, HAMD-28, HAMD-Item 1 (depressed mood), CGI severity, CGI improvement, HAMD-25 responders, CGI responders, and HAMD-25 remitters. Continuous variables were analyzed using the ANOVA model described above; CMH tests, adjusted for center, were used for categorical variables.
- *Comparisons of gepirone-ER vs. placebo and gepirone-ER vs. fluoxetine* — change from baseline for HAMD hypersomnia/hyperphagia factor (sum of items 22-26 of HAMD-31) and HAMA total score. ANOVA with effects for treatment (3 levels) and center was used to perform pairwise tests with an adjusted alpha of 0.025 for each comparison.

Results

Overall, 410 subjects were randomized and 409 received study medication (135 gepirone ER, 138 fluoxetine, and 136 placebo). A total of 391 subjects with post-baseline data comprised the ITT population (125 gepirone ER, 136 fluoxetine, and 130 placebo).

As seen in Table 44, 104 (25.4%) subjects did not complete the 8-week study. The drop-out rate was higher in the gepirone ER group (36.3%) than in the fluoxetine (18.1%) or placebo (21.3%) groups, mainly due to adverse events (10.4% vs. 2.9% and 1.5%) and "other reasons" (22.2% vs. 12.3% and 16.9%). These reasons included lost to follow-up, non-compliance, and withdrawn consent.

Table 44: Subjects Discontinued by Reason — Study 134004

Number of Subjects		Treatment Group			
		Gepirone ER	Placebo	Fluoxetine	Total
Randomized		135	137	138	410
Treated†		135	136	138	409
Discontinued	Total	49 (36.3%)	29 (21.3%)	25 (18.1%)	104 (25.4%)
	Adverse events	14 (10.4%)	2 (1.5%)	4 (2.9%)	20 (4.9%)
	Lack of efficacy	5 (3.7%)	4 (2.9%)	4 (2.9%)	13 (3.2%)
	Other reason	30 (22.2%)	23 (16.9%)	17 (12.3%)	70 (17.1%)
Completed Treatment		86 (63.7%)	107 (78.7%)	113 (81.9%)	306 (74.8%)

†One subject withdrew consent before the baseline visit and was randomized in error; no drug was dispensed.
[Source: CSR 134004 Tables 9 and 10; Appendix F Table 1.1.1 and 1.2.1]

Table 45 presents the change from baseline in the HAMD-25 total score by visit and treatment group.

The primary efficacy variable, the change from baseline in the HAMD-25 total score, showed no statistically significant differences between gepirone ER and placebo at any of the visits. There were no statistically significant differences between fluoxetine and placebo at any of the visits.

No treatment-by-center interactions were evident (p > 0.10) for any of these comparisons. Additionally, the OC results were consistent with LOCF analyses.

Table 45: HAMD-25 Total Score: Change from Baseline at Each Visit (ITT/LOCF) − Study 134004

Treatment Group		Baseline	LS Mean Change from Baseline					
			Week 1	Week 2	Week 3	Week 4	Week 6	Week 8/ET
gepirone ER (N=125)	N	125	124	124	124	124	124	124
	Mean	27.9	-3.72	-5.99	-7.99	-8.96	-9.61	-9.76
	SE	0.44	0.43	0.54	0.64	0.70	0.74	0.77
placebo† (N=130)	N	130	130	130	130	130	130	130
	Mean	27.6	-3.94	-6.75	-8.16	-10.12	-10.94	-10.63
	SE	0.44	0.42	0.53	0.63	0.68	0.72	0.75
gepirone ER vs. pbo	p-value		0.706	0.316	0.845	0.230	0.193	0.416
fluoxetine (N=136)	N	136	134	134	134	134	134	134
	Mean	28.1	-4.07	-5.77	-8.60	-9.93	-10.38	-11.66
	SE	0.46	0.43	0.53	0.60	0.64	0.73	0.75
placebo† (N=130)	N	130	130	130	130	130	130	130
	Mean	27.6	-3.84	-6.64	-8.12	-10.14	-10.91	-10.61
	SE	0.44	0.43	0.53	0.61	0.65	0.74	0.76
fluoxetine vs. pbo	p-value		0.699	0.242	0.576	0.817	0.605	0.325

ET = end of trial; LS = least squares; SE = standard error of the mean
†LS means and p-values from separate ANOVA models, with effects for treatment (2 groups: active drug vs. placebo) and center; preliminary tests of treatment x center interaction were not significant (p > 0.10) at any visit.
[Source: *CSR 134004 Table 16, Appendix F Tables 6.1.1.4 and 6.1.1.4A*]

Table 46 presents the results of the secondary efficacy endpoints for the ITT population (LOCF). Neither gepirone ER vs. placebo nor fluoxetine vs. placebo exhibited significant differences for any of the secondary efficacy variables.

Notably, the FDA performed an analysis of covariance (ANCOVA) (not the protocol specified analysis) on the HAMD-17 CFB (not the protocol specified endpoint). Even in this analysis, fluoxetine was not statistically better than placebo. The FDA compared fluoxetine to gepirone ER (not a protocol specified comparison). In this analysis, fluoxetine vs. gepirone ER results showed a mean difference of -1.71 (p = 0.027). The FDA concluded that this finding constituted assay sensitivity and judged the study to be negative rather than uninterpretable.

Table 46: Summary of Secondary Efficacy Results at Endpoint (ITT/LOCF) — Study 134004

Efficacy Variable	End of Treatment Outcome			Pairwise Tests (p-values)		
	Gep-ER	Fluoxetine	Placebo	G vs. P	F vs. P	G vs. F
HAMD-17 CFB	-5.67	-7.5	-6.55	0.282†	NR	NR
HAMD-28 CFB	-11.54	-14.0	-12.52	0.438	NR	NR
HAMD-Item1 CFB	-0.97	-1.2	-1.11	0.328	NR	NR
CGI (severity) CFB	-0.98	-1.2	-1.11	0.392	NR	NR
CGI (global improvement)	2.98	2.70	2.76	0.142	NR	NR
% Responders (HAMD-25)	33.87%	NR	36.15%	0.765	NR	NR
% Responders (CGI)	34.68%	NR	42.42%	0.224	NR	NR
% Remitters (HAMD-25)	16.94%	NR	23.85%	0.178	NR	NR
HAMD-31 (items 22-26) CFB	-2.82	-2.57	-2.80	0.948	NR	0.459
HAMD Factor 1 CFB (anxiety)	-2.09	-2.56	-1.93	0.594	NR	0.154
HAMA total score CFB	-4.08	-5.68	-4.95	0.226	NR	**0.025**

[*Source: 134004 CSR Appendix F Tables 6.1.1.1-6.1.2.8.1*]
LS means and p-values from ANOVA, with treatment (2 groups) and site as factors; Cochran-Mantel-Haenszel (CMH) test for % responders.
†Significant treatment-by-site interaction (p=0.05) for HAMD-17 CFB (gepirone ER vs. placebo).
NR = Not Reported; p-values are presented for all protocol-defined statistical comparisons.

Table 47 presents the change from baseline at for the HAMD-17 total score by visit and treatment group for the ITT population (LOCF).

For HAMD-17, differences between gepirone ER and placebo were not statistically significant at any visit. There is evidence of significant treatment by center interaction at Weeks 2, 3 and 8/ET, indicating that the effect of gepirone ER was inconsistent among sites. For this reason, findings for HAMD-17 based on the pooled sites must be interpreted cautiously. It is not clear if FDA performed tests of treatment by center interaction in the previously discussed ANCOVA model.

Table 47: HAMD-17 Total Score: Change from Baseline at Each Visit (ITT/LOCF) − Study 134004

Treatment Group		Baseline	Mean Change from Baseline					
			Week 1	Week 2	Week 3	Week 4	Week 6	Week 8/ET
gepirone ER (N=125)	**n**	125	124	124	124	124	124	124
	Mean	19.6	-1.5	-3.2	-4.6	-5.2	-5.5	-5.7
	SD	3.8	3.7	4.8	5.2	6.3	6.5	6.7
placebo (N=130)	**N**	130	130	130	130	130	130	130
	Mean	19.3	-1.8	-3.7	-4.7	-6.3	-6.8	-6.6
	SD	3.8	3.6	4.8	5.8	6.3	6.7	6.9
fluoxetine (N=136)	**n**	136	134	134	134	134	134	134
	Mean	19.9	-1.8	-2.9	-5.0	-6.1	-6.4	-7.5
	SD	4.1	3.6	4.4	5.3	5.5	6.3	6.7
gepirone ER vs. placebo		**p-value**	0.664	0.425	0.874	0.181	0.115	0.282
treatment x center interaction		**p-value**	0.415	**0.027**	**0.001**	0.162	0.223	**0.050**

ET = end of trial; SD = standard deviation
P-value based on protocol-defined ANOVA with effects for treatment (gepirone ER vs. placebo) and center.
[*Source: CSR 134004 Table 17, Appendix F Tables 6.1.2.1.1, 6.1.2.1.2 and 6.1.2.1.3*]

Conclusions

This was a 10-center, 3-arm, randomized, double-blind, active-controlled, study evaluating gepirone ER in comparison to placebo and fluoxetine over an 8-week treatment period in subjects diagnosed with MDD having "atypical features."

Since gepirone ER did not demonstrate a significant effect on the primary efficacy variable compared to placebo and neither did fluoxetine, based on the protocol, the study is uninterpretable.

FKP's advisors determined that Study ORG13 4004 is not a reliable basis for evaluating the efficacy of gepirone ER for the following reasons:

- *The study lacks assay sensitivity based on the primary endpoint (HAMD-25), which showed no statistically significant differences between treatment arms.*
- *HAMD-17 is not an appropriate endpoint for a study in atypical depression because it lacks items measuring symptoms typical of the illness and used as entry criteria.*
- *The comparator (fluoxetine) has not shown consistent efficacy in patients with MDD-AF. Clinical studies demonstrating its efficacy in MDD may have included a few patients with atypical depression, but did not select for such features, as in ORG134004.*
- *Subjects in ORG134004 (and ORG134006) are statistically different populations than the subject populations of all other gepirone-ER studies. They are also different populations from those used in registrational studies of all other antidepressants approved for the treatment of MDD.*

- *Lack of entry criteria for illness severity resulted in low severity of illness and wide variability within and between treatment groups, making comparisons uninterpretable.*
- *High placebo response reduced the chance to detect a treatment effect in this study.*
- *FDA's reliance on efficacy data from an analysis not specified in the protocol (ANCOVA) of a secondary endpoint (HAMD17) in this study violates ICH (E9 and E10) guidances, which warn against post-hoc analysis and use of an active comparator without consistently established efficacy in similar patient populations with atypical depression; also, use of comparative efficacy is not permitted under the FD&C Act.*
- *FDA ignored site interaction in the HAMD-17 analysis, casting doubt on the conclusion of fluoxetine's superiority to gepirone-ER.*
- *Numerical trends favoring placebo over gepirone-ER are likely due to biases of HAMD Q4-8 in atypical subjects, coupled with high variability of illness severity and high placebo response.*

9.1.6ORG134006

Study Description

This was a 13-center, 3-arm, randomized, double-blind, active-controlled study evaluating gepirone ER in comparison to placebo and paroxetine over an 8-week treatment period in subjects with atypical depression. The mean final dose of gepirone was 67.3 mg/day (±17.4), with 82.3% of subjects titrated to 60-80 mg/day. The mean final dose of paroxetine was 33.9 mg/day (±8.9), with 62.7% at the top dose.

Study Design

Randomized, double-blind, active-controlled, clinical study evaluating gepirone ER in the of atypical depression. The study was conducted at 13 sites, 12 in the US and 1 in Canada. 439 subjects were randomized and 437 were treated (147 gepirone-ER, 142 paroxetine, and 148 placebo). After a 7-20 day screening (placebo washout) period, eligible subjects were randomized to receive placebo, paroxetine (10-40 mg), or gepirone-ER (20-80 mg) once per day in the morning. The treatment period was to be 8 weeks with either a tapering period (5-15 days) or a 20-week double-blind extension phase. Study drugs were administered in a double-dummy, flexible dose manner. Dose titration was performed as follows: for gepirone-ER, the dose range was 20-80 mg/day with a single-blind forced titration step from 20 to 40 mg/day after 4 days of treatment; for paroxetine, the dose was increased from 10 to 20 mg/day after 4 days of treatment, then titrated between 20 and 40 mg/day after Day 8. Subjects were evaluated at baseline (Day 0) and Weeks 1, 2, 3, 4, 6 and 8.

Entrance Criteria

Inclusion Criteria

Key inclusion criteria were:

- Patients 18-65 years of age
- Patients who met criteria MDD (DSM-IV code 296) with either a single episode (296.2x) or recurrent episodes (296.3x)
- Patients who suffered from moderate depression (296.x2) to severe depression without psychotic features
- Patients who had no more than a 20% decrease on the HAMD-25 total score between screening and baseline
- Patients who met criteria for DSM-IV MDD with Atypical Features Specifier as assessed using the Atypical Depression Diagnostic Scale (ADDS)
- Patients who had dysphoria for most days over the past 4 weeks
- Patients who had a current episode of MDD with atypical features lasting at least 3 months

Efficacy Assessments

Primary Efficacy Measure

The primary efficacy endpoint was change from baseline (CFB) in HAMD-25 at week 8/ET (End of Treatment) for the ITT population using the LOCF approach. The HAMD-25 is the total score for items 1-18 and items 22-28 of the HAMD 31-item scale.

Secondary Efficacy Measures

Key secondary efficacy measures, were as follows:

- Comparisons of gepirone-ER vs. placebo: change from baseline on
  - HAMD-17
  - HAMD-28
  - HAMD-Item 1 (depressed mood)
  - Bech-6 core depression cluster
  - the number (%) of HAMD-25 responders and remitters
  - the number (%) of CGI responders
  - CGI improvement
  - CGI severity.
- Comparisons of gepirone-ER vs. placebo and gepirone-ER vs. paroxetine: change from baseline on
  - HAMD hypersomnia/hyperphagia factor (sum of items 22-26 of HAMD-31)

Statistical Methods

For the primary endpoint, the protocol-defined analysis was a two-group analysis of covariance (ANCOVA) model, with treatment (gepirone-ER vs. placebo) and center as factors, and baseline rating as a covariate. If the effect of gepirone-ER was not significant, paroxetine was compared to placebo (using a similar ANCOVA model) in order to validate the outcome. While not a traditional approach, this was the pre-planned analysis. The intent of the protocol was that if paroxetine was not statistically significantly different from placebo, there was no assay sensitivity, and further analysis was not warranted.

Secondary efficacy variables were analyzed as follows:

- *Comparisons of gepirone-ER vs. placebo* — change from baseline for HAMD-17, HAMD-28, HAMD-Item 1 (depressed mood), and Bech-6 core depression cluster, the number (%) of HAMD-25 responders and remitters, the number (%) of CGI responders, CGI improvement, and change from baseline in CGI severity. The ANCOVA model described above was used for continuous variables (ANOVA for CGI improvement scores); categorical variables were analyzed using CMH tests, stratified by center.
- *Comparisons of gepirone-ER vs. placebo and gepirone-ER vs. paroxetine* — change from baseline for HAMD hypersomnia/hyperphagia factor (sum of items 22-26 of HAMD-31). ANCOVA with effects for treatment (3 levels) and center was used to perform pairwise tests with an adjusted alpha of 0.025 for each comparison.

Results

Overall, 439 subjects were randomized and 437 were treated (147 gepirone-ER, 142 paroxetine, and 148 placebo). A total of 422 subjects with post-baseline data comprised the ITT population (143 gepirone-ER, 136 paroxetine, and 143 placebo).

As shown in Table 48, 125 (28.6%) subjects did not complete the 8-week treatment period. discontinuation rates were slightly higher in the gepirone-ER group (31.3%) than in the paroxetine (28.9%) or placebo (24.3%) groups, mainly due to adverse events (12.2% vs. 5.6% and 2.7%) and lack of efficacy (6.1% vs. 2.8% and 4.7%). Most drop-outs were attributed to "other reasons" (primarily lost to follow-up and withdrew consent): 12.9% gepirone-ER, 20.4% paroxetine, and 16.9% placebo.

Table 48: Subjects Discontinued by Reason — Study ORG134006

Number of Subjects		Treatment Group			
		Gepirone-ER	Placebo	Paroxetine	Total
Randomized		147	148	144	439
Treated†		147	148	142	437
Discontinued	Total	46 (31.3%)	36 (24.3%)	41 (28.9%)	125 (28.6%)
	Adverse events	18 (12.2%)	4 (2.7%)	8 (5.6%)	30 (6.9%)
	Lack of efficacy	9 (6.1%)	7 (4.7%	4 (2.8%)	20 (4.6%)
	Other reason	19 (12.9%)	25 (16.9%)	29 (20.4%)	73 (16.7%)
Completed Treatment		101 (68.7%)	112 (76.7%)	101 (71.1%)	314 (71.9%)

[*Source: CSR 134006 Tables 12 and 13; Appendix F Table 1.1.1 and 1.2.1*]

Table 49 presents the change from baseline in the HAMD-25 total score by visit and treatment for the ITT population (LOCF).

The primary efficacy variable, the change from baseline in HAMD-25 total score, showed no statistically significant differences between gepirone and placebo at any of the visits. No treatment by center interactions were evident ($p > 0.10$) for any of these comparisons.

There were also no statistically significant differences between paroxetine and placebo at any of the visits. Preliminary tests of treatment by center interaction were significant at Weeks 1, 2, 3, 6 and 8/ET, indicating that the effects of paroxetine were inconsistent among centers.

OC results were consistent with LOCF analyses, except that treatment by center interactions for the paroxetine vs. placebo comparisons were significant only at Weeks 1 and 2. Treatment effects were not statistically significant at any time point.

Table 49: HAMD-25 Total Score: Change from Baseline at Each Visit (ITT/LOCF) − Study ORG134006

Treatment Group		Baseline	LS Mean Change from Baseline					
			Week 1	Week 2	Week 3	Week 4	Week 6	Week 8/ET
gepirone-ER (N=143)	N	143	140	140	140	140	140	140
	Mean	27.0	-4.63	-6.24	-8.51	-9.00	-10.58	-10.94
	SE	0.38	0.48	0.53	0.57	0.63	0.69	0.74
placebo† (N=143)	N	143	143	143	143	143	143	143
	Mean	26.9	-4.87	-7.23	-8.37	-9.46	-10.63	-11.00
	SE	0.36	0.48	0.53	0.57	0.63	0.69	0.75
gepirone-ER vs. placebo	p-value		0.698	0.154	0.850	0.578	0.957	0.953
paroxetine (N=136)	N	136	136	136	136	136	136	136
	Mean	26.7	-4.95	-6.40	-9.18	-10.70	-12.13	-12.58
	SE	0.41	0.51	0.53	0.62	0.66	0.73	0.79
placebo† (N=143)	N	143	143	143	143	143	143	143
	Mean	26.9	-5.01	-7.47	-8.72	-9.88	-10.99	-11.23
	SE	0.36	0.49	0.51	0.60	0.64	0.70	0.76
paroxetine vs. placebo	p-value		0.917	0.114	0.566	0.326	0.220	0.178

ET = end of trial; LS = least squares; SE = standard error of the mean
† LS means and p-values from separate ANCOVA models, with effects for treatment (2 groups: active drug vs. placebo) and center; preliminary tests of treatment x center interaction were not significant ($p > 0.10$) at any visit for gepirone-ER vs. Placebo; the interactions were statistically significant for paroxetine vs. placebo at Weeks 1 (p=0.007), 2 (p=0.043), 3 (p=0.050), 6 (p=0.034) and 8/ET (p=0.024).
[Source: *CSR 134006 Table 19, Appendix F Tables 6.1.1.4A and 6.1.1.4AA*]

Table 50 presents the results of the secondary efficacy endpoints for the ITT population (LOCF).

There were no instances of statistically significant differences between gepirone-ER and placebo or between gepirone-ER and paroxetine for any of the protocol-defined comparisons, except for HAMD-31 Items 22-25 (hyperphagia/hypersomnia) which show significantly greater reductions in the gepirone-ER group compared to the paroxetine group at Week 3 (p=0.011) and Week 8/ET (p=0.012).

The final report for this study, in accordance with the original analysis plan, did not provide Based on all planned comparisons of secondary efficacy variables, there were no statistically significant treatment effects for either gepirone-ER or paroxetine.

Table 50: Summary of Secondary Efficacy Results at Endpoint (ITT/LOCF) — Study ORG134006

Efficacy Variable	End of Treatment Outcome			Pairwise Tests (p-values)*		
	Gep-ER	Paroxetine	Placebo	G vs. P	Px vs. P	G vs. Px
HAMD-17 CFB	-6.92	-9.1	-7.15	0.750	NR	NR
HAMD-28 CFB	-12.68	-14.8	-12.57	0.927	NR	NR
HAMD-Item1 CFB	-1.11	-1.4	-1.07	0.720	NR	NR
CGI (severity) CFB	-1.10	-1.4	-1.21	0.423	NR	NR
CGI (global improvement)	2.68	2.3	2.63	0.726	NR	NR
HAMD Bech-6 score CFB	-4.60	-6.0	-4.55	0.918	NR	NR
% Responders (HAMD-25)	42.86%	NR	41.96%	0.894	NR	NR
% Responders (CGI)	45.71%	NR	46.85%	0.808	NR	NR
% Remitters (HAMD-25)	30.71%	NR	32.87%	0.652	NR	NR
HAMD-31 (items 22-26) CFB	-2.76	-2.10	-2.47	0.253	NR	**0.012**

[*Source: ORG134006 Final Report Appendix F Tables 6.1.1.4AA-6.1.2.3.3*]
*LS means and p-values from ANCOVA, with treatment (gepirone-ER vs. placebo) and center as factors, baseline value as covariate; ANCOVA applied to 3 treatments for HAMD-31 (items 22-26) only. Cochran-Mantel-Haenszel (CMH) test used for % responders.
NR=Not Reported; p-values are presented for all protocol-defined statistical comparisons.

Table 51 presents the results of the FDA's Re-analysis.

The FDA performed an ANCOVA on the HAMD-17 CFB (not the protocol-specified primary endpoint). Presumably, FDA applied the model to all 3 groups to make pairwise treatment comparisons using a common error term (not as specified in the protocol).

Based on HAMD-17 CFB, paroxetine was significantly better than placebo (p=0.026). The FDA also compared paroxetine to gepirone-ER (not a protocol specified comparison). In this analysis, paroxetine showed a better effect than gepirone-ER (mean difference -1.85; p=0.012). The FDA concluded that this constituted assay sensitivity and judged the study to be negative rather than uninterpretable.

Table 51: FDA Findings for HAMD-17 CFB at Endpoint — Study ORG134006

Treatment Comparison	Effect Size	p-value
gepirone-ER vs placebo	0.22	0.760
paroxetine vs placebo	-1.63	**0.026**
gepirone-ER vs paroxetine	-1.85	**0.012**

Conclusions

This was a 13-center, 3-arm, randomized, double-blind, active-controlled study evaluating gepirone ER in comparison to placebo and paroxetine over an 8-week treatment period in subjects with atypical depression.

The intent of the study's protocol was such that, if gepirone-ER was not statistically different from placebo and paroxetine was also not statistically different from placebo, the study was uninterpretable.

Since gepirone-ER did not demonstrate a significant effect on the primary efficacy variable compared to placebo and neither did paroxetine, based on the protocol, the study is uninterpretable.

FKP's advisors determined that Study ORG13 4006 is not a valid basis for evaluating the efficacy of gepirone ER for the following reasons:

- *The study lacks assay sensitivity based on the primary endpoint, HAMD-25, the proper measure of symptoms in MDD-AF, which showed no statistically significant treatment effects for paroxetine or gepirone-ER.*
- *FDA determined that this is a negative trial based on results for HAMD-17, not the primary efficacy variable. HAMD-17 is not an appropriate endpoint for a study in atypical depression because it lacks items measuring symptoms typical of the illness and used as entry criteria.*
- *The comparator (paroxetine) has not shown consistent efficacy in patients with MDD-AF. Clinical studies demonstrating its efficacy in MDD may have included a few patients with atypical depression, but did not select for such features, as in ORG134006.*
- *Subjects in study ORG134006 (and ORG 134004) are statistically different populations than the subject populations of other gepirone-ER studies. They are also different populations from those used in registrational studies of all other antidepressants approved for the treatment of MDD.*
- *Lack of entry criteria for illness severity resulted in wide variability within and between treatment groups, making comparisons uninterpretable.*
- *Contributing to the failure of this trial was the 46% placebo response rate (by CGI) and the fact that 55% of subjects enrolled did not meet minimum criteria for depression. High placebo response reduced the chance to detect a treatment effect in this study.*
- *FDA's reliance on efficacy data from this study violates ICH (E9 and E10) guidances, which warn against post-hoc analysis and use of an active comparator without established efficacy in similar patient populations with MDD-AF; also, use of comparative efficacy is not permitted under the FD&C Act.*
- *A significant treatment by site interaction for HAMD-25 CFB indicates that the direction of paroxetine's effect is inconsistent among the study sites; similar differences were noted among sites in HAMD-17 CFB, casting further doubt on the use of this measure as a basis for judging assay sensitivity in this trial.*

9.1.7 134017

Study Description

This was a 9-center, randomized, double-blind, active-controlled study evaluating gepirone ER (20-80 mg/day) in comparison to placebo and fluoxetine over an 8-week treatment period subjects with MDD (165 gepirone ER, 166 fluoxetine, and 165 placebo). The average dose for

gepirone ER was 58.7±15.0 mg/day, with 78.8% of subjects receiving a final prescribed dose in the range of 60-80 mg/day. The average dose of fluoxetine was 25.9±4.7 mg/day, with 72.7% at the top dose.

Study Design

Randomized, double-blind, active-controlled, clinical study evaluating gepirone-ER in the treatment of MDD. The study was conducted at 9 sites, in the US and randomized 496 subjects. After a 7-day placebo wash-out period, eligible subjects were randomized to receive either placebo, gepirone-ER tablets (20-80 mg), or fluoxetine (20-40 mg) once per day in the morning for 8 weeks (56 days). Study drugs were administered in a double-dummy, flexible dose manner. Dose titration was performed as follows: for gepirone-ER, the dose range was 20-80 mg/day with a single-blind forced titration step from 20 to 40 mg/day after 4 days of treatment; for fluoxetine, the dose could be increased from 20 to 40 mg/day after the first 4 weeks of treatment. Subjects were evaluated at baseline (Day 0) and Weeks 1, 2, 3, 4, 6 and 8.

Entrance Criteria

Inclusion Criteria

Key inclusion criteria were:

- Patients 18-65 years old
- Patients with a primary diagnosis of recurrent MDD (DSM-IV 296.3)
- Patients with a baseline HAMD-17 total score ≥ 18 (amended to 22 during study)
- Patients with dysphoria for most days over the past 4 weeks
- Patients with normal sexual function and been sexually active in the 12 months prior to baseline

Efficacy Assessments

Primary Efficacy Measure

The primary efficacy endpoint was the change from baseline (CFB) in MADRS total score at Week 8 or end of treatment (ET) using the LOCF approach.

Secondary Efficacy Measures

Secondary efficacy measures, were as follows:

- Change from baseline on
  - HAMD-17
  - HAMD-25
  - HAMD-Item 1 (depressed mood),
  - CGI-severity score

- CGI-improvement score
- HAMD-17 responders
- GI responders
- HAMD-17 remitter
- HAMD-Factor 1 (hypersomnia and hyperphagia).

None of these secondary parameters was designated as "key" in the protocol.

Statistical Methods

For the primary endpoint, the protocol specified use of an ANOVA model, with treatment (3 groups) and site as factors, to perform all pairwise treatment comparisons. A preliminary test of treatment-by-center interaction was also performed; this term was dropped from the model if $p > 0.10$.

Fluoxetine was included in the study design to evaluate assay sensitivity. A comparison of fluoxetine to placebo with respect to change from baseline MADRS total score was performed at the 0.05 level of significance within the same ANOVA model described above. The intent of the protocol was that if fluoxetine was not statistically significantly different from placebo there was no assay sensitivity, and further analysis was not warranted.

The comparison between gepirone-ER and fluoxetine was considered a secondary analysis. This comparison was made by testing at the 0.05 level of significance the null hypothesis that there was no difference between gepirone-ER and fluoxetine with respect to the change from baseline MADRS total score. This was performed within the same ANOVA model as described above.

For the secondary endpoints, continuous variables were analyzed via an ANOVA with treatment and center as factors; pairwise tests of each active drug to placebo were planned if the overall test of treatment effect was statistically significant at the .05 level. For categorical variables, analyses were performed using CMH test, adjusting for center.

Results

Overall, 496 subjects were randomized and 495 were treated (165 gepirone-ER, 166 fluoxetine, and 164 placebo). A total of 480 subjects had post-baseline efficacy assessments and comprised the ITT population (160 gepirone-ER, 161 fluoxetine, and 159 placebo).

As seen in Table 52, the drop-out rate was higher in the gepirone-ER group (31.5%) than in the fluoxetine (24.1%) or placebo (21.3%) groups, mainly due to adverse events (8.5% vs. 4.8% and 1.2%) and "other reasons" (17.6% vs. 16.9% and 15.9%). The majority of subjects discontinued for reasons not specified on the CRF. These reasons included lost to follow-up, non-compliance, and withdrawn consent.

Table 52: Subjects Discontinued by Reason - Study ORG134017

Number of Subjects		Treatment Group			
		Gepirone-ER	Placebo	Fluoxetine	Total
Randomized		165	165	166	496
Treated†		165	164	166	495
Discontinued	Total	52 (31.5%)	35 (21.3%)	40 (24.1%)	127 (25.7%)
	Adverse events	14 (8.5%)	2 (1.2%)	8 (4.8%)	24 (4.8%)
	Lack of efficacy	9 (5.5%)	7 (4.3%)	4 (2.4%)	20 (4.0%)
	Other reason	29 (17.6%)	26 (15.9%)	28 (16.9%)	83 (16.8%)
Completed Treatment		113 (68.5%)	129 (78.7%)	126 (75.9%)	368 (74.3%)

[*Source: CSR 134017 Tables 8 and 9; Appendix F Table 1.1-3 and 1.2-1*]

Table 53 presents the change from baseline in the MADRS score by visit and treatment group for the ITT population (LOCF).

No statistically significant differences in the change from baseline in the MADRS total score were found between the gepirone-ER group and the placebo group at Week8/ET ($p = 0.650$) or at any other visit. The comparison between fluoxetine and placebo also failed to show statistical significance at endpoint ($p = 0.299$) and at other visits. A possible explanation may be the high placebo response, which was larger than expected. A secondary comparison of gepirone-ER versus fluoxetine showed no statistically significant difference between the two active treatment groups in the change from baseline of the MADRS total score at endpoint ($p = 0.136$) or any other visit.

Notably, The FDA has also argued that there are too many instances where placebo or the comparator drug is better than gepirone-ER. As shown above, on the primary efficacy parameter MADRS CFB, gepirone-ER is 0.5 points less than placebo and fluoxetine is better than both.

Table 53 MADRS: Change from Baseline at Each Visit (ITT/LOCF) — Study ORG134017

Treatment Group		Baseline	LS Mean Change from Baseline					
			Week 1	Week 2	Week 3	Week 4	Week 6	Week 8/ET
gepirone-ER (N=160)	**n**	159	159	159	159	159	159	159
	Mean	29.5	-3.86	-7.78	-9.46	-11.56	-12.29	-12.23
	SE	0.37	0.50	0.61	0.70	0.74	0.80	0.84
placebo (N=159)	**n**	159	159	159	159	159	159	159
	Mean	29.6	-4.94	-7.35	-9.60	-11.23	-12.47	-12.73
	SE	0.41	0.50	0.61	0.69	0.73	0.79	0.84
fluoxetine (N=161)	**n**	159	159	159	159	159	159	159
	Mean	29.1	-4.59	-7.54	-9.66	-12.27	-12.56	-13.88
	SE	0.41	0.50	0.61	0.70	0.74	0.80	0.84
Treatment (Overall)		p-value	0.244	0.865	0.975	0.547	0.965	0.310
gep-ER vs. placebo		p-value	0.100	0.591	0.879	0.730	0.860	0.650
fluoxetine vs. placebo		p-value	0.600	0.809	0.947	0.282	0.934	0.299
gep-ER vs. fluoxetine		p-value	0.263	0.768	0.827	0.465	0.795	0.136

LS means and p-values from ANOVA model, with treatment and site as main effects; treatment x site interactions were not significant (p > 0.10).
ET = end of trial; LS = least squares; SE = standard error of the mean
[*Source: CRS ORG134017 Table 15, Appendix F Table 6.1-1.3*]

Table 54 presents the results of the secondary efficacy endpoints for the ITT population (LOCF).

For HAMD-17 CFB, the overall test of treatment effect was not statistically significant at end of study (p=0.137) or at any other time point except Week 1 (p=0.025); at Week 1, reductions in HAMD-17 were numerically greater in the placebo group than in either of the active treatment groups.

No statistically significant differences were observed between gepirone-ER and placebo groups in the number (%) of HAMD-17 Responders and Remitters. At Week8/Endpoint, HAMD-17 Responder and Remitter rates were significantly higher for fluoxetine than gepirone-ER group (p=0.006 and p=0.044, respectively); differences between gepirone-ER and fluoxetine were not statistically significant at any other study visits. Note that, at Week 8/Endpoint, the number of HAMD-17 responders in the placebo group was 72/159 (45%).

No significant treatment effects were noted at Week 8/Endpoint (or any other study visit) for HAMD-25 CFB, HAMD-Item 1 CFB, CGI severity CFB, CGI improvement scores, or the number (%) of CGI Responders. At Week 8/Endpoint (and all other study visits), the HAMD hyperphagia score (items 25-26) was unaffected by gepirone-ER and placebo, but showed significant reductions in the fluoxetine group compared to gepirone-ER.

Notably, The FDA did a post hoc ANOVA (not a protocol-specified analysis) on HAMD-17 (not the primary efficacy endpoint) to compare fluoxetine vs. gepirone-ER (not a protocol specified comparison), which resulted in a difference favoring fluoxetine of -1.54, p=0.042.

Table 54: Summary of Secondary Efficacy Results at Endpoint (ITT/LOCF) — Study ORG134017

Efficacy Variable	End of Treatment Outcome			Pairwise Tests (p-values)*		
	Gep-ER (N=160)	Fluoxetine (N=161)	Placebo (N=159)	G vs. P	F vs. P	G vs. F
HAMD-17 CFB	-10.23	-11.76	-10.96	--	--	--
HAMD-25 CFB	-12.03	-13.77	-12.86	--	--	--
HAMD-Item1 CFB	-1.33	-1.50	-1.24	--	--	--
CGI (severity) CFB	-1.24	-1.41	-1.32	--	--	--
CGI (global improvement)	2.48	2.29	2.44	--	--	--
% Responders (HAMD-17)	42.14%	57.23%	45.28%	0.607	NR	**0.006**
% Responders (CGI)	54.09%	62.26%	52.83%	0.829	NR	0.152
% Remitters (HAMD-17)	22.01%	32.08%	31.45%	**0.060**	NR	**0.044**
HAMD (items 22-24) hypersomnia	-0.59	-0.43	-0.65	--	--	--
HAMD (items 25-26) hyperphagia	-0.10	-0.38	-0.06	0.742	NR	**0.037**

[*Source: CSR ORG134017 Tables 16-25*]
*P-values and LS means from ANOVA for continuous variables, with treatment and center as the main effects; -- indicates that a pairwise test was not performed because the overall treatment effect was not statistically significant. CFB=Change from Baseline.
CMH test was used for % responders (gepirone-ER vs each other group). NR = Not Reported; this was not a protocol-defined statistical comparison.

Conclusions

This was a 9-center, randomized, double-blind, active-controlled study evaluating gepirone ER (20-80 mg/day) in comparison to placebo and fluoxetine over an 8-week treatment period subjects with MDD.

The intent of the protocol was such that if fluoxetine was not statistically significantly different from placebo for the primary endpoint there was no assay sensitivity, and further analysis was not warranted, rendering the study uninterpretable.

Since fluoxetine was not statistically significantly different from placebo for the primary endpoint, based on the protocol, the study is uninterpretable.

FKP's advisors concluded that study ORG134017 is not interpretable for the following reasons:

- *Study ORG134017 has no assay sensitivity because it failed to demonstrate a significant effect of the active control, fluoxetine, on the primary endpoint (MADRS).*
- *The failure of this study to detect treatment effects can be attributed to the placebo response rate (53% based on CGI), which was too high for the study to succeed.*
- *The failure of this study to detect treatment effects is likely also impacted by enrollment of a significant percentage of subjects without severe enough levels of depression; the entry criterion was based on HAMD-17 scores rather than the primary efficacy variable (MADRS), and modified during the enrollment period; 22% of subjects had baseline HAMD-17 < 22.*

- *The study was poorly executed, with significant protocol compliance issues and highly inconsistent results among sites: out of 9 sites, 5 favored gepirone-ER and fluoxetine over placebo and 4 favored placebo over gepirone-ER and fluoxetine.*

- *The observed non-significant results in HAMD-17 favoring placebo over gepirone-ER are not meaningful in light of the inordinately high placebo response, inconsistent results among sites, variability introduced by protocol non-compliance, and the effect of drop-outs in the LOCF analysis.*

10. APPENDIX 3 - SUMMARY OF RELAPSE PREVENTION STUDY (28709)

APPENDIX 3

SUMMARY OF RELAPSE PREVENTION STUDY

Study Description

Study 28709 was a European, multicenter, placebo-controlled trial in MDD outpatients. The trial started with an open-label (OL) gepirone ER treatment phase of 8-12 weeks. Responders, defined as those reaching a HAMD-17 of 8 or less, were entered into a double blind continuation phase of 40-44 weeks.

Study Design

Eligible subjects with MDD were screened and treated with open-label gepirone ER (20-80 mg/day). Visits during the open-label (OL) phase were at weeks 1, 2, 4, 8 and 12. Responders to gepirone ER (HAMD-17 of 8 or less) within the 8-12 week window were randomized to the double-blind continuation period of 40-44 weeks, with a visit scheduled every 4 weeks. Relapse was defined by a HAMD-17 score of 16 or more, or discontinuation due to ineffectiveness of study drug. Subjects who relapsed were to be withdrawn from the trial. For randomized subjects, the OL phase was the period from the date of first dose of trial medication up to and including the date of randomization. The continuation phase started the date after randomization and ended on the date of last dose of study medication.

The Intent-to-Treat population (ITT) consisted of all randomized subjects who received study medication and had at least one efficacy assessment during the double-blind continuation phase. The primary efficacy analysis was based on the ITT population. The primary efficacy parameter was the number (%) of subjects with a relapse at the end of the continuation phase, with relapse defined by:

- HAMD-17 total score ≥ 16, or
- Discontinuation due to 'Relapse Criteria Fulfilled' on the End of Trial (EOT) form.

The primary time-point for treatment comparisons was the endpoint assessment of the continuation phase based on the ITT group.

Statistical tests of these variables (ANOVA for continuous data, CMH or Wilcoxon tests for categorical data, adjusting for center) were performed for exploratory reasons only.

Eligible subjects were 18-70 years of age with a primary diagnosis of recurrent major depressive disorder (DSM-IV 296.3), had a HAMD-17 total score of at least 20 at screening and baseline.

Primary Efficacy Measure

The primary objective of this trial was to compare the relapse rates of depression during the continuation phase between subjects receiving gepirone ER at the final titrated dose and subjects receiving placebo.

Secondary Efficacy Measures

- Compare time to relapse during the continuation phase between subjects receiving gepirone ER at the final titrated dose and subjects receiving placebo;

- Evaluate the therapeutic efficacy and the safety profile of gepirone ER at the final titrated dose during the continuation phase in comparison with placebo;
- Investigate the effects of gepirone ER at the final titrated dose during the continuation phase on sexual functioning and quality of life in comparison with placebo.

Original Results

In total, 435 subjects were screened, 428 were selected to participate in the OL phase, and 420 were actually treated with gepirone ER. Of the 303 subjects who completed the OL phase, 250 subjects met responder criteria and were randomized into the double-blind continuation phase and treated with trial medication: 126 gepirone ER, 124 placebo. The mean modal doses of gepirone during the OL and continuation phases were 60 mg and 80 mg. During the continuation phase, 51 of the 126 subjects were exposed to daily doses of 60 mg or more for at least another 36 weeks. The mean daily dose in the OL phase was 57.1±14.1 mg/day. In the double-blind phase the mean dose was 61.9±17 mg/day.

At the end of the OL phase, 68.7% of the subjects had improved after treatment with gepirone ER according to the global impression of change; and 61.4% of the subjects had a HAMD-17 total score ≤ 8, indicating that the subjects had responded to gepirone ER treatment.

Table 55 presents relapse rates at each study visit during the double-blind continuation phase (ITT/LOCF).

Table 55: Relapse Rate by Week of Double-Blind Continuation Phase (ITT/LOCF) – Study 28709

Assessment	% of Subjects with a Relapse		p-value
	gepirone ER N=126	placebo N=124	
Week 4	8.7%	6.5%	0.804
Week 8	11.1%	9.7%	0.826
Week 12	12.7%	12.1%	0.812
Week 16	13.5%	16.9%	0.345
Week 20	15.1%	21.8%	**0.079**
Week 24	15.9%	24.2%	**0.034**
Week 28	16.7%	29.0%	**0.009**
Week 32	17.5%	31.5%	**0.011**
Week 36	19.8%	33.1%	**0.011**
Week 40	23.0%	34.7%	**0.024**
Week 44	23.0%	34.7%	**0.024**

[*Source: CSR Appendix F Table 6.2.1-A.2 and Analysis 6.2.1-A.3*]
P-value from CMH test adjusted for center.

The relapse rate at endpoint of the continuation phase (based on the primary analysis of the primary efficacy parameter) was 23.0% for subjects in the gepirone ER group compared to 34.7% for subjects in the placebo group (p=0.024). Beginning at Week 16, a difference in relapse rates became apparent; from Week 24 onward, the differences were statistically significant in favor of gepirone ER. When 5 subjects who discontinued the study for other reasons were

included as relapses, trends were similar but differences were significant only at Week 28 (p=0.032).

The difference between treatment groups in time to relapse did not achieve statistical significance based on the log-rank test (p=0.065).

Table 32 presents results for secondary efficacy endpoints (change from baseline in HAMD-Item 1, HAMD Response, and CGI Response) at the end of the double-blind continuation phase (Week 44) based on the ITT/LOCF approach.

Table 32: Summary of Secondary Efficacy Results at End of Double-Blind Continuation Phase (ITT/LOCF) — Study ORG28709

Parameter	End of Continuation Phase		p-value
	gepirone ER N=124	placebo N=123	
Compared to Baseline=Prior to Start of OL Treatment Phase:			
HAMD-Item 1 CFB	-1.9	-1.6	**0.065**
HAMD-17 CFB	-15.3	-13.4	**0.083**
% with HAMD-17 ≤ 8	64.5%	51.6%	**0.059**
HAMD-17 Responders (%)	71.8%	64.2%	0.215
HAMD-21 Responders (%)	72.6%	65.0%	0.178
HAMD-25 Responders (%)	74.2%	65.9%	**0.095**
HAMD-28 Responders (%)	74.2%	65.0%	**0.082**
CGI Responders (%)	71.0%	63.7%	0.137
CGI Impression of Change	-2.5	-2.1	0.122
Compared to Baseline=Randomization (Prior to Start of DB Continuation Phase):			
Global Impression Improved (%)	15.3%	15.3%	0.755
Global Impression Worsened (%)	22.6%	32.2%	0.121

[*Source: ORG28709 CSR Table 25, Table 26, Table 27 and Appendix F Analysis 6.2.2-A.2, 6.2.2-K.2, 6.2.2-M.2 and 6.6.2-M.5*]
CFB=Change from baseline (as defined in table); LS means from ANOVA model, including effects for treatment and center.
HAMD responders are subjects with ≥ 50% reduction from baseline; p-values from CMH test, controlling for center
CGI responders are much or very much improved on the CGI improvement score; p-value is based on Wilcoxon test of CGI scores, adjusting for center.

HAMD-Item 1 (Depressed Mood**):** From Week 20 onward, a (borderline) statistically significant difference in mean change from baseline was observed for the HAMD-Item 1 indicating more sustained improvement in depressed mood with gepirone ER than placebo. At endpoint, the mean change from baseline was -1.9 for the gepirone ER group and -1.6 for the placebo group (p=0.065).

HAMD Responders: For both treatment groups, the percentage of subjects maintaining response on the HAMD-17 total score decreased during the continuation phase. From Week 12 onward, the percentage was higher in the gepirone ER group than in the placebo group. This difference favoring gepirone ER was significant from week 20 to week 32. A similar pattern was evident for responder rates based on HAMD-21, HAMD-25, and HAMD-28. At all timepoints, the

percentage of subjects with HAMD-17 ≤ 8 was higher for the gepirone ER group than the placebo group.

CGI Responders: From Week 12 of the continuation phase onward, the percentage of subjects who were "much improved" or "very much improved" was higher in the gepirone ER group as compared to placebo.

Previous Assessments

Evaluation by FKP prior to FDA Submission by Organon

Prior to FDA submission, FKP evaluated the report prepared by Organon and found that 5 subjects that were clear gepirone ER relapses had not been counted. In addition, 32 subjects had been deleted from the database because they came from centers with only 1 treatment arm represented, or with no relapses occurring. Tables 31 and 32 above reflected these inappropriate exclusions. FKP recognized these problems and admonished Organon not to submit this report to the FDA.

FDA Re-analysis and FKP Response

The FDA noted the same problems uncovered by FKP. Adding in the deleted subjects and relapses, the FDA concluded that gepirone ER was not effective in this trial, since the difference in relapse rates did not achieve statistical significance: gepirone ER group 27% and placebo group 32.7% (p=0.10).

The FDA interpreted the failure of 28709 to achieve statistical significance as an indication that gepirone ER did not have adequate efficacy in depression. FDK disagreed with this negative interpretation, primarily because of the numerous problems of study design and implementation.

Notably FKP conducted an analysis with subjects who relapse immediately after randomization as not true responders.

This analysis demonstrates that the difference in relapse rates among true responders is statistically significant favoring gepirone ER. The numbers of such occurrences make the results very suspicious: 11 subjects meet these criteria (8 gepirone ER and 3 placebo subjects). All were relapses in the ITT analysis. If these non-responders are removed from the calculations, relapse rates are: gepirone ER 22/126 (17.5%) and placebo 40/124 (32.3%), with p = 0.007. The above subjects were classified as relapses on their first visit after randomization, but only 2 had values of 8 or less for more than 1 visit prior to randomization, both in the placebo group (257 and 636). Adding these subjects back into the analysis, the relapse rates are: gepirone ER 22/126 (17.5%) and placebo 42/124 (33.9%), with p = 0.003. The FDA advised FKP that a HAMD-17 score of 8 or less is not necessary for randomization in a placebo substitution study, that a 50% drop in baseline HAMD-17 is adequate to define remission. If we include subjects with a 50% drop in HAMD-17 for more than one visit, the results include 5 more subjects: 3 gepirone ER subjects and 2 placebo subjects. The re-calculated relapse rates are: 25/126 gepirone ER (19.8%) vs.

42/124 placebo (33.9%), with p = 0.013. Table 56 summarizes results based on the various alternative definitions for remission.

Table 56: Relapse Rates in Study 28709 (Alternate Analysis)

Randomized	No. (%) Relapse		p-value*
	gepirone-ER N = 126	placebo N = 124	
Original ITT	29/126 (23.0%)	43/124 (34.7%)	**0.024**
Corrected ITT	34/126 (27.0%)	43/124 (34.7%)	0.101
Per Protocol	25/104 (24.0%)	41/106 (38.7%)	**0.023**
Excluding non-responders[1]	22/126 (17.5%)	40/124 (32.3%)	**0.007**
Excluding non-responders[2]	22/126 (17.5%	42/124 (33.9%)	**0.003**
Including 50% drop in HAMD[3]	25/126 (19.8%)	42/124 (33.9%)	**0.013**

*Chi-square test of proportions (two-sided); relapse=HAMD-17 ≥ 16
[1] Excludes relapses on 1st visit after randomization.
[2] Excludes relapses on 1st visit after randomization if response was confirmed prior to randomization
[3] Includes subjects with 50% drop in HAMD-17 prior to randomization as responders.

Current Assessment

FKP has recently reviewed the evidence from Study 28709. This study was far from perfectly implemented, with various protocol and study conduct limitations. However, even the corrected analysis by FDA provides some support for efficacy, as there is a clear trend favoring gepirone ER. Moreover, there is no reason to believe that these problems would create a bias in favor of gepirone ER. To the extent that errors were effectively random, the study's power might have been reduced, but the results remain somewhat supportive of efficacy, with a clear trend favoring gepirone ER.

11. APPENDIX 4 – GEPIRONE ER STATISTICAL REVIEW (JANUARY 2011)

APPENDIX 4

Gepirone ER Statistical Review (January 2011)

STATISTICAL REVIEW AND EVALUATION

NDA: 21-164

Drug Name: Gepirone Hydrochloride Extended-Release (ER)

Indication: Major Depressive Disorder

Sponsor: Fabre-Kramer Pharmaceuticals

Documents Reviewed:

Final Study Reports and Appendices (12 Controlled Clinical Trials)
Integrated Efficacy Summary dated April 2007
FDA Non-approvable letter dated November 2, 2007
Fabre Kramer response to FDA dated December 21, 2007
FDA official minutes of meeting dated January 14, 2008

Statistical Reviewer: Mary F. Johnson, PhD
Exec VP, Biostatistics
PharmaNet, Inc.

Contents

Tables

Figures

Background

This review pertains to efficacy results from clinical studies evaluating gepirone ER for the treatment of major depressive disorder (MDD). The review was undertaken at the request of Fabre Kramer (FK), as they wish to obtain an independent assessment of studies comprising the NDA for their product. FK sought my input because of my background as a statistical reviewer in the Biometrics Division at FDA and my experience designing/analyzing clinical trials in compliance with FDA regulations and guidelines.

The NDA for gepirone ER was originally filed by Organon, Inc. in July 2001. Non-approvable letters were issued by FDA on March 15, 2002 and June 23, 2004. In the second letter, FDA noted that one study (134001) demonstrated short-term efficacy in MDD, but that a second confirmatory study would be required. The NDA was subsequently transferred to Fabre Kramer (FK). On May 1, 2007, FK re-submitted the NDA with a complete response to the June 23, 2004 non-approvable letter, including results from a new study (FKGBE007). FDA issued a non-approvable letter dated November 2, 2007 indicating that 007 was pivotal, but enumerating several reasons why the 12 clinical studies conducted to date failed to substantiate efficacy. In a follow-up meeting on January 14, 2008, FDA further elaborated their concerns and doubts about the persuasiveness of evidence in support of the product's efficacy.

My critique of the studies focuses on the quality and statistical validity of efficacy results from each study, as a basis for judging the "weight of evidence" in support of gepirone's efficacy. During the review, I had access to final clinical study reports, NDA documents, and study-related correspondence. FK compensated my time for this work, but any conclusions reached regarding individual studies or the program as a whole were reached independently based on my experience and review of the available data.

Controlled Clinical Studies of Gepirone ER

The NDA includes results from a total of 12 randomized, controlled studies evaluating gepirone ER for the treatment of MDD. In general, the studies were of short-term treatment duration (6-8 weeks) and efficacy variables included well-recognized clinical measures of depression, including the Hamilton Depression Rating Scale (HAMD), the Clinical Global Impression Scale (CGI), and Montgomery-Asberg Depression Rating Scale (MADRS). The change from baseline in HAMD-17 total score was designated as the primary efficacy variable in most of the studies.

The results presented in individual study reports complied with methods and outcome variables pre-specified in each protocol. Typically, treatment groups were compared using analysis of variance (ANOVA) applied to change from baseline values of each rating scale, adjusting for center effects, and Cochran-Mantel-Haenszel (CMH) tests for categorical data. The Integrated Summary of Efficacy (ISE) also provided results of analysis of covariance (ANCOVA) models, with factors for treatment, center, and baseline value as covariate. Results were reported for each study visit using both OC (observed case) and LOCF (last observation carried forward) methods to account for drop-outs. This review will focus on LOCF results in the ITT (intent-to-treat) population (all subjects randomized, treated, and having post-baseline data), unless otherwise stated.

Relevant details of each study are summarized below, and cross-referenced with information in the Final Reports. ANCOVA results and additional analyses from the ISE are also cited, as needed. The studies are presented in chronological order according to dates of completion.

Placebo-Controlled Studies (No Active Comparator)

Study CN105-078 [Dec 1991-Aug 1992]

This study was a 2-center, randomized, double-blind, placebo-controlled flexible dose trial evaluating gepirone ER at 2 dose levels (10-50 mg/day [low dose] and 20-100 mg/day [high dose]) in subjects with MDD (baseline HAMD-17 $\geq$ 20) during a 6-week treatment period. By protocol, the primary efficacy analysis was a comparison of change from baseline HAMD-17 total score and the percentage of CGI responders for the pooled gepirone dose groups vs. placebo, whereas the secondary analysis compared each individual gepirone ER dose group with placebo.

The study was designed to enroll 180 subjects (120 in the combined gepirone ER groups and 60 in the placebo group), which would provide at least 72% power to detect a 3-point difference in change from baseline HAMD-17 total score between gepirone ER and placebo at the 5% significance level (assuming SD=7.4). However, the study was terminated early when Bristol Myers Squibb (BMS) decided to stop the product's development for business reasons after 144 subjects (80%) had started double-blind treatment, thus reducing the power of the study to approximately 62%.

Of the 144 subjects randomized and treated (50 low dose, 45 high dose, 49 placebo), a total of 5 subjects (3.5%) were discontinued due to early study termination and 55 subjects (38.2%) failed to complete the 6-week treatment period. Reasons for drop-out included adverse events (10.0% low dose, 31.1% high dose, and 8.2% placebo), lack of efficacy (2.0% low dose, 0.0% high dose, and 10.2% placebo), subject unreliability (4.0% low dose, 6.7% high dose, and 2.0% placebo) and other reasons such as lost to follow-up and withdrew consent (14% low dose, 20.0% high dose, and 10.2% placebo). A total of 135 subjects with post-baseline data comprised the ITT population (48 low dose, 40 high dose, 47 placebo).

The mean maximum dose of gepirone ER administered during this study was 37.3 ±11.2 mg/day in the gepirone ER low-dose group, compared to 67.5±22.0 mg/day in the gepirone ER high-dose group. Thus, half of subjects in the gepirone ER groups received maximum doses below the minimum effective dose (40 mg/day).

The primary efficacy analysis, which evaluated HAMD-17 change from baseline in the combined gepirone ER dose groups (ITT/LOCF), showed no significant drug-placebo differences at any study visit. Results of the ITT/OC analysis showed statistically significant differences in favor of the high-dose gepirone ER and total gepirone ER treatment groups over placebo based on the change from baseline in HAMD-17 total score at Week 4. In addition, the percentage of CGI responders was significantly greater in the high-dose gepirone group than in the placebo group at Week 4.

Table 1 summarizes results for all primary and secondary efficacy variables at the final visit of this study.

Table 1: Efficacy Results in Study CN105-078

Efficacy Variable	End of Treatment Outcome (ITT/LOCF)			Pairwise Tests (p-values)	
	Gep Low	Gep High	Placebo	Low vs. P	High vs. P
	N=48	N=40	N=47		
HAMD-25 CFB	-9.8	-10.1	-8.7	0.492	0.406
HAMD-17 CFB	-7.5	-7.5	-6.5	0.460	0.473
HAMD-28 CFB	-11.2	-11.8	-10.2	0.546	0.403
HAMD-Item1 CFB	-1.0	-1.1	-0.7	0.169	0.082
CGI (severity) CFB	-0.8	-1.0	-0.7	0.630	0.186
CGI (global improvement)	2.8	**2.5**	**3.0**	0.386	**0.041**
% Responders (HAMD-17)	33%	35%	28%	0.533	0.465
% Responders (CGI)	50%	55%	38%	0.227	0.128
MADRS CFB	-8.7	-9.9	-7.1	0.457	0.239
HAMD Factor 1 CFB (anxiety/somatization)	-2.2	-2.0	-1.9	0.513	0.733
HAMD Factor V CFB	-2.9	**-3.5**	**-2.1**	0.122	**0.017**
HAMD Factor VI CFB	-0.9	-0.7	-1.3	0.282	0.130
SCL-87 total score CFB	-28.9	-42.9	-39.7	0.381	0.809
SCL-87 depression CFB	-8.7	-13.4	-9.4	0.815	0.210

Source: CN105-078 Final Report Appendices F7.1, 7.2, and 7.4
LS means/p-values from ANOVA (treatment and center as factors); CMH test for % responders
HAMD-17 Responder = 50% improvement
CGI Responder = Much improved or very much improved

Endpoint results favored high dose gepirone over placebo for all variables, with significant differences detected for CGI global improvement (p=0.041) and HAMD Factor V (p=0.017).

The sponsor performed post-hoc mixed model analyses for HAMD-17 and several other parameters, including the modified HAMD-17 (mHAMD-17), the HAMD core depression factor (Bech-6), HAMD Item 1, and the MADRS total score. Mixed-models analyses were performed under 2 different covariance structures: first-order autoregressive structure and compound symmetry for the ITT dataset. Based on either model, the high dose group showed significantly greater improvement than the placebo group in mHAMD-17 ($p \leq 0.012$), Bech-6 ($p \leq 0.032$), and HAMD-Item 1 ($p \leq 0.039$) scores. [Note: The mHAMD-17 scale replaces items related to insomnia, nausea and agitation with ratings of neuro-vegetative symptoms from the HAMD-25. I could find no reference for this scale, so it appears to be an ad hoc modification.].

Comment: This study had limited potential to detect treatment effects, given that it was planned with low power (72%), terminated early, and employed relatively low doses of gepirone in half of the subjects. An interim analysis prepared for BMS at the time of early termination (Makuch, 1992) noted that the observed treatment differences were consistent with the effect size used in the original sample size calculations but that, due to inadequate power, no conclusions could be reached. Compounding the problem, a high percentage of subjects (38%) dropped out of this study during the 6-week treatment period (34% low dose, 51% high dose, and 31% placebo),

mainly due to adverse events, lack of efficacy, and loss to follow-up/non-compliance. While the study is flawed, trends suggest that depression symptoms improved to a greater extent in the high dose gepirone group than in the placebo group. The sponsor's re-analysis of the modified HAMD-17 score and other secondary variables is not persuasive, given that it was not pre-specified in the analysis plan and exploits trends in the data observed after unblinding. This statistical approach seems to be a futile attempt to derive useful information from a study that was not adequate to reach its stated objectives. Overall, based on deficiencies in both design and execution, the study should be designated a failure, providing no useful data for assessing the efficacy of gepirone ER.

Study CN105-083 [Dec 1991-Aug 1992]

This was a 2-center, randomized, double-blind, placebo-controlled, flexible dose trial evaluating 2 doses of gepirone ER (10-50 mg/day [low dose] and 20-100 mg/day [high dose]) in subjects with MDD (baseline HAMD-17 score ≥ 20) during a 6-week treatment period. By protocol, the primary efficacy analysis was a comparison of change from baseline HAMD-17 total score and the percentage of CGI responders for the pooled gepirone dose groups vs. placebo, whereas the secondary analysis compared each individual gepirone ER dose group with placebo.

The originally planned sample size for this study was 180 subjects (120 in the combined gepirone ER groups and 60 in the placebo group), which would provide at least 72% power to detect a 3-point difference in the mean change from baseline HAMD-17 total score between gepirone ER and placebo at the 5% significance level. However, the study was stopped early by BMS (for administrative reasons) after only 117 subjects (65%) had enrolled, reducing the power of the study to approximately 53% to detect the originally hypothesized treatment effect.

Of the 117 subjects randomized and treated (37 low dose, 39 high dose, 41 placebo), a total of 10 subjects (8.5%) were discontinued prematurely when the study was abruptly stopped and 43 subjects (36.8%) failed to complete the 6-week treatment period. Reasons for drop-out included adverse events (13.5% low dose, 12.8% high dose, and 12.2% placebo) and a variety of other reasons such as lost to follow-up and withdrew consent. Most of the drop-outs (34 of 43 or 79%) occurred within the first 3 weeks of treatment. A total of 112 subjects with post-baseline data comprised the ITT population (36 low dose, 37 high dose, 39 placebo)

The mean maximum dose of gepirone ER administered during this study was 37.2 mg/day in the low dose group (below the minimum effective dose of 40 mg/day) and 70.3 mg/day in the high dose group.

No statistically significant treatment effects were detected for the primary efficacy variable based on the ITT LOCF population. As specified in the protocol, results for the primary endpoint were assessed for the presence of a treatment-by-center interaction (defined as $p \leq 0.10$). For the primary analysis (ITT/LOCF), a significant treatment-by-center interaction was noted in the model comparing the pooled gepirone treatment groups to placebo (p=0.098). This interaction was found to be qualitative, with gepirone and placebo response rates notably different at each of the two sites.

At site 001, the gepirone subjects scored somewhat lower than at site 002 (adjusted mean changes from baseline -8.6 vs. -10.3, respectively). Additionally, the placebo response appeared

greater in site 001 than in site 002 (adjusted mean change from baseline -10.7 vs. -7.2, respectively). Finally, site 001 included more subjects in the ITT dataset (n=63) than in site 002 (n=47). Consequently, it is reasonable to consider the sites independently. For study site 002, the high dose gepirone ER-treated evaluable subject group was statistically significant for both the CGI responder and HAMD 17 change from baseline observed cases analyses (Table 2). Study site 001 did not show statistically significant results for any efficacy parameter.

Table 2: Summary of Efficacy Variables at Endpoint by Treatment Group, Site 002 ITT and Evaluable Group (LOCF Analysis, Study CN105-083)

Parameters	LS Mean Change from Baseline to Endpoint and Percent Responders					
	Gepirone ER 10-50 mg/day		Gepirone ER 20-100 mg/day		Placebo	
	ITT	OC	ITT	OC	ITT	OC
Primary Efficacy Variables						
HAMD-17	-10.2	-10.7	-10.4	-12.5[a]	-7.2	-7.1
CGI Responders[c]	59%	60%	69%[b]	85%[a]	38%	36%
Secondary Efficacy Variables						
HAMD-28	-14.0	-14.8	-14.8	-18.0[a]	-9.0	-9.3
HAMD Item 1	-1.4	-1.5	-1.3	-1.5	-0.9	-1.0
HAMD-17 Responders[d]	53%	53%	50%	62%[b]	31%	29%
MADRS	-9.6	-9.8	-14.3	16.3[b]	-9.8	-9.7

Source: CN105-083 Final Report Appendix F

[a] p-value ≤ 0.05 vs. placebo
[b] $0.05 < p \leq 0.10$
[c] Responder defined as a subject who is much or very much improved based on the CGI improvement score at any post-baseline assessment.
[d] Responder defined as a subject with $\geq$ 50% reduction in baseline HAMD-17 total score at study endpoint.

Comment: This study had the same limitations as Study CN105-078: reduced power and sample size due to early study termination, a high drop-out rate (36.8%) possibly due to rapid dose escalation over the 6-week treatment period, and under-dosing (less than the minimum effective dose) in the low dose gepirone ER group. Once again, no statistically significant drug-placebo differences were detected for the primary efficacy variable in ITT LOCF populations. An interim analysis report (Makuch, 1992) advised BMS that this study (similar to CN105-078) was severely under-powered, such that no reliable conclusions could be reached. The study is difficult to interpret and inadequate as a basis for efficacy conclusions because of its relatively small sample size, low range of gepirone doses employed, high drop-out rate in the first 3 weeks (partly due to early termination), and evidence of treatment-by-center interaction. Data from one of the two study sites (site 002) suggest that high dose gepirone ER had activity in the OC dataset, but this finding for a secondary variable in a small subset is likely to be spurious. In short, study CN105-083 is inconclusive regarding the efficacy of gepirone ER.

Study 134001 [Jun 1999 - Dec 2000]

This was a 5-center, randomized, double-blind, placebo-controlled study evaluating gepirone ER 20-80 mg/day over an 8-week treatment period in subjects with MDD. The study had one

forced-titration step from 20-40 mg/day at Day 4, and a flexible-dose design thereafter. The mean dose (± SD) of gepirone ER was 61.05 (±12.02) mg/day.

Overall, 208 subjects were randomized and treated (102 gepirone ER, 106 placebo). The drop-out rate was slightly higher in the gepirone group (27.5% vs. 23.6%), mostly due to adverse events (9.8% vs. 2.8%); other reasons, including lost to follow-up or withdrawn consent, were more frequent in the placebo group (13.7% vs. 17.0%). Four subjects in each group dropped out for lack of efficacy. A total of 204 subjects with post-baseline data comprised the ITT population (101 gepirone ER, 103 placebo).

Statistically significant differences between the gepirone ER and placebo treatment groups were noted for the change from baseline in the HAMD-17 total score at Week 3 and at Week 8/Endpoint (p=0.013) based on ANCOVA with baseline as covariate. Also, marginally significant differences in favor of gepirone ER were observed for the change from baseline in the HAMD-17 total score at Weeks 1, 2, 4 and 6 ($p < 0.10$; 134001 Final Report, Appendix F8.6.1.1-4). Results based on the ITT/Observed Cases (OC) analysis were consistent with these findings.

Table 3 presents results for HAMD-17 at final visit (ITT/LOCF) for each center and overall based on the ANCOVA model, with factors for treatment, center, and baseline value as covariate. Center-specific results for HAMD-17 showed trends favoring gepirone in 4 of the 5 centers; the treatment effect in center 1 achieved statistical significance on its own (p=0.011). Across centers, the average reduction in HAMD-17 was significantly greater in gepirone-treated subjects than in the placebo group (-9.04 vs -6.57; p=0.013). The treatment by center interaction term was not statistically significant (p=0.385).

Table 3: Analysis of Change in HAMD-17 from Baseline to End of Double-Blind Treatment Period in Study 134001

	Number of subjects		(Adjusted) mean change		Treatment difference and 95% CI		
Center	Gepirone ER	Placebo	Gepirone ER	Placebo	(Gepirone ER - Placebo)	SE	p-value
Center 1	33	32	-10.46	-6.59	-3.88 (-6.82, -0.94)	1.47	0.011
Center 2	24	25	-7.84	-8.71	0.87 (-2.90, 4.63)	1.87	0.645
Center 3	25	24	-11.58	-8.31	-3.27 (-7.68, 1.15)	2.19	0.143
Center 4	15	15	-9.57	-5.03	-4.54 (-11.44, 2.37)	3.36	0.189
Center 5	4	5	-6.03	-3.77	-2.26 (-10.73, 6.22)	3.46	0.539
All Centers Combined	101	101	-9.04	-6.57	-2.47 (-4.41, -0.53)	0.98	**0.013**
Treatment by center interaction							0.385

Note: ANCOVA model was used with terms for treatment and center and baseline value (as a covariate).
Source: Statistical Table 3.1

The results of this study showed statistically significant differences in favor of gepirone ER for the primary efficacy variable (change from baseline in HAMD-17 total score), as well as several secondary variables (change from baseline in total scores for HAMD-21, HAMD-25, HAMD-28, CGI severity and HAMD-Item 1, HAMD-25 responders, and HAMD-17 remitters). A summary of efficacy results in Study 134001 is presented in Table 4.

Table 4: Summary of Primary and Secondary Efficacy Parameters (Study 134001)

Parameter	Gepirone ER Change from Baseline mean±SE	Placebo Change from Baseline mean±SE	Difference	p-value at Endpoint
HAMD-17 total score[a]	-9.04 ± 0.78	-6.75 ± 0.77	-2.47	0.018
HAMD-21 total score[a]	-10.01 ± 0.88	-7.49 ± 0.87	-2.54	0.021
HAMD-25 total score[a]	-11.57 ± 1.01	-8.19 ± 0.99	-3.38	0.007
HAMD-28 total score[a]	-13.27 ± 1.20	-9.60 ± 1.18	-3.67	0.013
%HAMD-17 Responders[b]	43.6%	30.7%	12.9%	0.059
%HAMD-25 Responders[c]	45.5%	28.7%	16.8%	0.014
%HAMD-17 Remitters[d]	28.7%	14.9%	13.8%	0.017
CGI severity[a]	-1.19 ± 0.13	-0.79 ± 0.13	-0.4	0.015
HAMD Item-1[a]	-1.16 ± 0.11	-0.78 ± 0.11	0.38	0.005

Source: 134001 Final Report Appendix F

[a] ANOVA model does not incorporate baseline value as a covariate.
[b] Responder defined as a subject with ≥ 50% reduction in baseline HAMD-17 total score at any post-baseline assessment.
[c] Responder defined as a subject with ≥ 50% reduction in baseline HAMD-25 total score at any post-baseline assessment.
[d] Remitter defined as a subject with a HAMD-17 total score of ≤ 7 at any post-baseline assessment.

Comment: The treatment effect was statistically significant for all primary and secondary efficacy variables presented in Table 4 ($p \leq 0.021$), except for the HAMD-17 responder analysis which showed borderline significance (p=0.059). The study appears to be adequately designed and executed, employing doses in the appropriate therapeutic range for gepirone ER. Results for the primary efficacy variable were statistically significant, consistent among investigators, and supported by secondary efficacy variables. Thus, Study 134001 provides valid evidence of the effectiveness of gepirone ER in reducing symptoms of MDD.

Study 134002 [Jun 1999 – Dec 2000]

This was a 5-center, randomized, double-blind, placebo-controlled study of gepirone ER 20-80 mg/day over an 8-week treatment period in subjects with MDD. The study had a flexible-dose design; the minimum final dose of gepirone ER was 40 mg/day. The mean dose (± SD) of gepirone ER was 57.90 (± 13.03) mg/day. The final prescribed dose was 60 and 80 mg/day in 23.4% and 58.9% of subjects, respectively.

Overall, 218 subjects were randomized and treated (110 gepirone ER, 108 placebo). The drop-out rate was slightly higher in the gepirone group (31.8% vs. 28.7%), mostly due to adverse

events (10.0% vs. 7.4%) and other reasons (19.1% vs. 18.5%) such as lost to follow-up and withdrew consent. Three subjects in each group dropped out for lack of efficacy. A total of 211 subjects with post-baseline data comprised the ITT population (107 gepirone ER, 104 placebo).

This study failed to show a statistically significant treatment effect for the primary efficacy variable (change from baseline in HAMD-17 total score) at endpoint (see Table 5) or any other study visit; trends in mean values directionally favored gepirone ER over placebo consistently from Week 1 to endpoint.

Among secondary efficacy variables, HAMD-Item 1 (depressed mood) showed significantly greater improvement in the gepirone group compared to placebo at Weeks 2, 4, 6 and end of trial (-1.30 vs -1.01, p=0.036). Results based on the OC analysis showed statistically significant differences in the HAMD-17 responders at Week 8 (p=0.044); and marginally significant differences favoring gepirone ER at Weeks 4 and 6 ($0.05 < p \leq 0.10$).

In general, the study showed trends favoring gepirone ER over placebo and statistical significance at some time points for some parameters, but not for the primary efficacy parameter at endpoint.

Drop-out rates were more common on gepirone ER than placebo and occurred somewhat earlier (18% vs. 14% within the first 2 weeks of the study), but these findings do not explain the lack of positive efficacy findings. Placebo response rates were relatively high in this trial (32% based on the HAMD-17 and 45% based on the CGI).

Table 5: Summary of Efficacy Results at Endpoint, ITT/LOCF (Study 134002)

Parameter	Gepirone ER	Placebo	Difference	P-Value at Endpoint
HAMD-17 CFB	-9.96 ± 0.65	-9.29 ± 0.65	-0.67	0.446
Bech-6 CFB	-5.90 ± 0.53	-4.95 ± 0.53	-0.95	0.076
HAMD Item 1 CFB	-1.30 ± 0.10	-1.01 ± 0.10	-0.18	0.036
MADRS CFB	-11.52± 0.97	-9.17 ± 0.91	-2.34	0.078
				CMH p-value
CGI Responders [a]	52.0%	44.7%	7.3%	0.297
HAMD-17 Responders [b]	40.2%	32.0%	8.2%	0.225

Source: 134002 Final Report Appendix F; 2003 Summary of Benefits and Risks Appendix A-5.
CFB=Change from baseline to endpoint; CMH=Cochran Mantel-Haenszel.
[a] Responder defined as a subject who is much or very much improved based on the CGI improvement score at any post-baseline assessment.
[b] Responder defined as a subject with ≥ 50% reduction in baseline HAMD-17 total score at any post-baseline assessment.

The sponsor performed post-hoc mixed model analyses (as described above for Study CN105-078), for HAMD-17 and several other parameters, including the modified HAMD-17 (mHAMD-17), the HAMD core depression factor (Bech-6), HAMD Item 1, and the MADRS total score. Table 6 presents results of the mixed-models analyses under 2 different covariance structures: first-order autoregressive structure (Model 1) and compound symmetry (Model 2) for the ITT dataset. By taking advantage of repeated measurements, the mixed model approach

demonstrated significant treatment effects favoring gepirone ER over placebo for each of the secondary efficacy variables. Results are strongly positive for the mHAMD-17, Bech-6, Item-1, and MADRS in both analyses.

Table 6: Post-Hoc Mixed Models Analyses Results (Study 134002)

Parameter	Gepirone ER Change from Baseline	Placebo Change from Baseline	Mixed Models p-value
HAMD-17			
Mixed Model 1	-8.81 ± 0.42	-7.97 ± 0.41	0.135
Mixed Model 2	-8.93 ± 0.42	-8.08 ± 0.41	0.137
mHAMD-17 [a]			
Mixed Model 1	-8.27 ± 0.41	-6.67 ± 0.51	0.004
Mixed Model 2	-8.33 ± 0.41	-6.75 ± 0.41	0.005
Bech-6			
Mixed Model 1	-5.25 ± 0.26	-4.06 ± 0.25	0.001
Mixed Model 2	-5.30 ± 0.26	-4.10 ± .025	0.001
HAMD Item 1			
Mixed Model 1	-1.20 ± 0.07	-0.89 ± 0.06	0.001
Mixed Model 2	-1.21 ± 0.07	-0.90 ± 0.06	0.001
MADRS			
Mixed Model 1	-10.63 ± 0.61	-8.09 ± 0.60	0.002
Mixed Model 2	-10.71 ± 0.61	-8.18 ± 0.60	0.002

Source: 2003 Summary of Benefits and Risks, Appendix A-6.

[a] Modified HAMD -17 is formed by replacing the insomnia and appetite items (Insomnia early, Insomnia middle, Insomnia late, Somatic symptoms gastrointestinal, and Loss of weight) with HAMD-25 items that measure the opposite (or reverse) neurovegetative symptoms (Hypersomnia: Time in bed, Oversleeping, and Napping; Increased appetite, Weight gain). These item substitutions provide a scale that removes the effect of 5 items that may be less sensitive in some subjects who are receiving compounds that can produce insomnia, nausea, and agitation, but that includes relevant reverse neurovegetative symptoms frequent in MDD.

Comment: Given that Study 134002 did not achieve its primary objective, the study provides rather weak evidence of efficacy based on post-hoc analyses and limited significant findings for secondary efficacy variables. Mixed models analysis is a powerful statistical tool that takes advantage of repeated measures over time, but its use was data-driven, not pre-specified prior to blind-break, and clearly undertaken to strengthen the case for efficacy. The study was adequate in design and properly conducted to meet its stated objectives, but results are only weakly supportive of the efficacy of gepirone ER in MMD.

Study 134023 [May 2003 – Mar 2004]

This was a 12-center, randomized, double-blind, placebo-controlled, flexible dose trial evaluating gepirone ER in subjects with MDD during an 8-week treatment period. The mean

dose (± SD) of gepirone ER was 61.3 (±13.66) mg/day, with 69.3% of subjects reaching a final dose of 80 mg/day.

Overall, 254 subjects were randomized and treated (127 per group). The drop-out rate was higher in the gepirone ER group than in the placebo group (26.0% vs. 21.3%), with more subjects discontinuing gepirone for adverse events (9.4% vs. 0.8%), but fewer for lack of efficacy (3.9% vs. 5.5%). Most of the drop-outs were due to "other" reasons (12.6% vs. 15.0%), including lost to follow-up, withdrawn consent, and protocol non-compliance. A total of 246 subjects (123 per group) with post-baseline data comprised the ITT population.

No statistically significant treatment effects were detected for gepirone ER based on the primary or secondary efficacy variables. HAMD-17 scores showed an average reduction of 8 points in both treatment groups at the end of study (p=0.947); baseline means were comparable (approximately 23) in the two groups.

Comment: This study appears to be adequately designed and powered to detect treatment effects in an appropriate population of subjects diagnosed with MDD. There is no clear explanation for the negative results. A relatively large number of investigators participated in the study and recruited subjects over a short (10 month) period. This may have contributed to study failure (e.g., heterogeneity, less rigorous subject selection criteria, or lack of standardization in riteria or symptom assessments), but this is only conjecture.

Study FK-GBE-007 [Oct 2003 – Aug 2004]

This was a 9-center, randomized, double-blind, placebo-controlled study evaluating gepirone ER 20-80 mg/day over an 8-week treatment period in subjects with MDD. The study had one forced dose titration step from 20 mg/day to 40 mg/day between Days 4 and 7, and a flexible dose design thereafter. The mean dose (± SD) of gepirone ER was 58.2 (±13.95) mg/day, with 87.9% of subjects reaching a final dose of 60-80 mg/day.

Overall, 248 subjects were randomized (124 gepirone ER, 124 placebo). The drop-out rate was higher for gepirone ER (21.8% vs. 17.7%), with slightly more discontinuing treatment due to lack of efficacy (3.2% vs. 2.4%), adverse events (4.0% vs. 2.4%), and other reasons (14.5% vs. 12.9%) such as lost to follow-up and withdrew consent. Ten randomized subjects (8 gepirone, 2 placebo) were excluded from the ITT population, as they had no post-baseline assessments within 3 days of study drug administration.

The results of this study showed statistically significant differences in favor of gepirone ER for the primary efficacy variable (change from baseline HAMD-17 total score using LOCF), and nearly all the secondary efficacy variables: CGI responders, HAMD-17 responders, HAMD-17 remitters, MADRS, change from baseline in total scores for the HAMD-21, HAMD-25, HAMD-28, and the HAMD item 1 (depressed mood).

Table 7 presents results for HAMD-17 based on the ANCOVA model, with factors for treatment, center, and baseline value as covariate; Table 8 presents results of the ANOVA model, without baseline adjustment.

Based on the ITT/LOCF population, the treatment effect was statistically significant for the change from baseline HAMD-17 total score at Weeks 4, 6 and 8. The difference was also marginally significant at Week 3 (p=0.081; see Study FK-GBE-007 Final Report Table 16.) Results based on the ITT/OC analysis were consistent with these findings.

Overall, the average reduction in HAMD-17 was significantly greater in gepirone-treated subjects than in the placebo group (-10.24 vs -7.79; p=0.018). The treatment by center interaction was nearly significant (p=0.092), indicating some inconsistency in results among centers. Center-specific results for HAMD-17 are displayed in Figure 1. Reductions in HAMD-17 were greater in the gepirone group compared to the placebo group in 5 of 8 centers; the treatment effect achieved statistical significance in 2 of these centers (centers 701 and 706, with p=0.005 and p=0.044, respectively). Trends favored placebo in 3 centers: center 999 (pooled centers 703 and 707, that each enrolled fewer than 16 subjects), center 704 and center 705; none of these differences achieved statistical significance individually.

Table 7: Analysis of Change in HAMD-17 from Baseline to End of Double-Blind Treatment Period (Study FK-GBE-007)

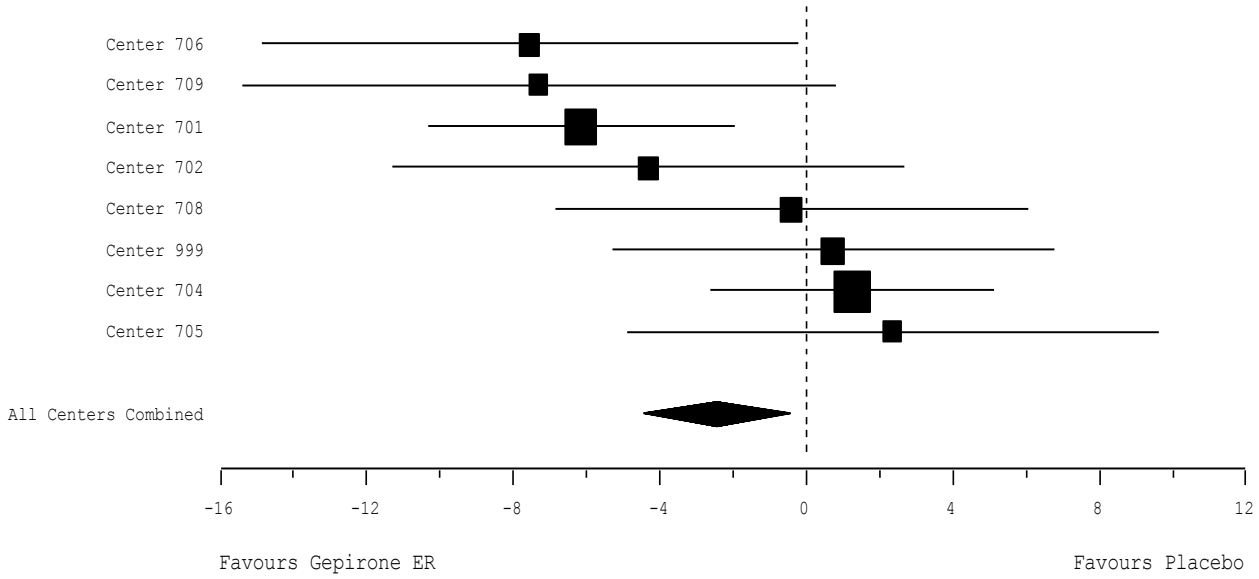
Center	Number of subjects		(Adjusted) mean change		Treatment difference and 95% CI	SE	p-value
	Gepirone ER	Placebo	Gepirone ER	Placebo	(Gepirone ER - Placebo)		
Center 701	21	23	-11.49	-5.34	-6.15 (-10.33, -1.97)	2.07	0.005
Center 702	14	16	-12.33	-8.02	-4.31 (-11.30, 2.67)	3.40	0.216
Center 704	22	20	-10.38	-11.63	1.25 (-2.63, 5.13)	1.92	0.519
Center 705	14	15	-5.98	-8.35	2.36 (-4.88, 9.61)	3.53	0.509
Center 706	10	9	-12.89	-5.34	-7.55 (-14.85, -0.24)	3.45	0.044
Center 708	12	10	-10.91	-10.51	-0.40 (-6.85, 6.05)	3.08	0.898
Center 709	8	11	-11.55	-4.24	-7.31 (-15.41, 0.79)	3.82	0.074
Center 999	15	18	-7.20	-7.94	0.74 (-5.28, 6.75)	2.95	0.804
All Centers Combined	116	122	-10.24	-7.79	-2.45 (-4.47, -0.43)	1.02	**0**.018
Treatment by center interaction							0.092

Note: ANCOVA model was used with terms for treatment and center and baseline value (as a covariate).

Center 999 is a pooled center that combines centers 703 and 707.

Source: ISE Statistical Table 3.2

Figure 1: Forest Plot for the Change in HAMD-17 from Baseline to End of Double-Blind Treatment Period (Study FK-GBE-007) BY CENTER



Source: ISE Statistical Figure 1.2.

A summary of efficacy results in Study FK-GBE-007 is shown in Table 8.

Table 8: Summary of Primary and Secondary Efficacy Parameters (Study FK-GBE-007)

Parameter mean±SE	Gepirone ER Change from Baseline	Placebo Change from Baseline	Difference	p-value at Endpoint
HAMD-17 total score [a]	-10.22 ± 0.75	-7.96 ± 0.73	-2.26	0.032
HAMD-21 total score [a]	-11.07 ± 0.80	-8.79 ± 0.78	-2.28	0.043
HAMD-25 total score [a]	-12.65 ± 0.91	-9.85 ± 0.89	-2.80	0.029
HAMD-28 total score [a]	-15.04 ± 1.06	-11.83 ± 1.04	-3.21	0.032
CGI severity [a]	-1.30 ± 0.11	-0.92 ± 0.11	-0.38	0.015
HAMD Item 1 [a]	-1.39 ± 0.12	-1.07 ± 0.12	-0.32	0.056
MADRS total score [a]	-13.72 ± 1.01	-9.94 ± 0.99	-3.28	0.008
%HAMD-17 Remitters	34.5%	20.5%	14%	0.019

Source: FK-GBE-007 Final Report Supportive Tables 19- 21, 27-29, 41-43, 49-51, 57-59, 65-67, 71-73, and 103.

[a] ANOVA model, without incorporating baseline value as a covariate.

Responder rates are shown in Table 9. Statistically significant results were obtained for all of these variables.

Table 9: Responder Rates at Week 8/Endpoint (Study FK GBE-007)

Responder Rate at Week 8/ET (ITT, LOCF)					
	Gepirone ER		Placebo		
Variable	n/N	%	n/N	%	p-value
HAMD-17 [a]	53/116	45.7	36/122	29.5	0.014
MADRS [a]	59/116	50.9	39/121	32.2	0.005
HAMD-21 [a]	54/116	46.6	39/122	32.0	0.031
HAMD-25 [a]	56/116	48.3	37/122	30.3	0.007
HAMD-28 [a]	57/116	49.1	40/122	32.8	0.015
CGI [b]	56/116	48.3	42/121	34.7	0.045

Source: FK-GBE-007 Final Report, Supportive Tables 22, 30, 44, 52, 60 and 74.

[a] Responder defined as a subject with ≥ 50% reduction in baseline HAMD-17, HAMD-25, or MADRS total score at any post-baseline assessment.
[b] Responder defined as a subject who is much or very much improved based on the CGI improvement score at any post-baseline assessment.

Comment: As specified in the protocol for study FK-GBE-007, any significant treatment-by-center interactions (defined as a p-value ≤ 0.10) would be further explored. For the primary efficacy variable (HAMD-17) as reported above, a significant p-value of 0.092 was noted.

A full analysis of the implications of the treatment-by-center interaction was prepared by a statistical consultant (Dr. Gene Laska); see Appendix D of the ISE. He found significant center effects at baseline and Weeks 2, 3 and 4, but not at Weeks 6 and 8. The effect of gepirone ER was statistically superior to placebo at Weeks 4, 6 and 8 and in some analyses at Week 3. Several sensitivity analyses were performed (e.g., pooling centers as per the SAP, not pooling centers, and dropping extreme centers) and all led to the same conclusion regarding the treatment effect. That is, the choice of pooling methodology did not alter conclusions of the study with respect to significant treatment effects. In particular, dropping the best and worst centers eliminated the significance of the interactions at all visits and had no effect on the significance of the treatment effects. Taken together, these analyses provide a convincing case that the treatment effect is not diminished by the effect of pooling centers or the presence of the interaction term.

Study FK-GBE-008 [Oct 2003 – Aug 2004]

This was an 8-center randomized, double-blind, placebo-controlled, flexible dose study evaluating gepirone ER 20-80 mg/day over an 8-week treatment period in subjects with MDD. [The study is identical in design to Study FK-GBE-007, except for a smaller planned sample size: 100 subjects/group instead of 120.] The mean dose of gepirone ER was 60.0±13.1 mg/day. By the final visit, 86.9% of subjects were at a dose of 60-80 mg/day.

Overall, 206 subjects were randomized (102 gepirone ER, 104 placebo). The drop-out rate was slightly higher in the gepirone ER group (24.5% vs. 21.1%), primarily due to AEs (4.9% vs.

1.9%) and lost to follow-up (13.7% vs. 8.7%); drop-outs for other (reasons such as lack of efficacy, withdrawn consent, and non-compliance) occurred with low and comparable frequency in the 2 groups.

P-values resulting from statistical tests of treatment effect for primary and secondary efficacy variables at each study visit and endpoint are shown in Table 10.

Table 10: Statistical Significance of Efficacy Results at Each Time Point, ITT/LOCF (Study FK-GBE-008)

	Week 2	Week 3	Week 4	Week 6	Week 8/ET
HAMD-17 CFB	0.016†	0.053	0.123	0.046†	0.159
MADRS CFB	0.008†	0.039†	0.035†	0.008†	0.208
CGI Severity	0.101	0.214	0.075	0.070	0.273
HAMD-21 CFB	0.023†	0.147	0.170	0.055	0.209
HAMD-25 CFB	0.057	0.135	0.243	0.087	0.281
HAMD-28 CFB	0.025†	0.212	0.308	0.119	0.319
HAMD Item 1 CFB	0.035†	0.321	0.649	0.552	0.469
CGI Responders [a]	0.132	0.011†	0.100	0.037†	0.147
HAMD-17 Responders [b]	0.096	0.176	0.287	0.050	0.293
HAMD-21 Responders [b]	0.054	0.552	0.458	0.068	0.231
HAMD-25 Responders [b]	0.288	0.676	0.215	0.024†	0.035†

† Statistically significant. CFB=Change from Baseline
Source: FK-GBE-008 Final Report, Tables 21-22, 29, 43-44, 51-52, 59, 67, 73, and 74.
[a] Responder defined as a subject who is much or very much improved based on the CGI improvement score at any post-baseline assessment.
[b] Responder defined as a subject with ≥ 50% reduction in baseline HAMD total score at any post-baseline assessment.

This study failed to demonstrate a statistically significant treatment effect for the protocol-defined primary endpoint (change from baseline in HAMD-17 scores at Week 8). However, trends in mean values directionally favored gepirone ER over placebo at each visit, with significant differences detected at Week 2 and Week 6 (p=0.016 and p=0.046, respectively). By Week 8, mean scores were 14.3 and 15.7 (p=0.159).

Comment: The sponsor tried to reconcile HAMD-17 findings in this study with more positive results in study FK-GBE-007, noting similar effect sizes but standard deviations that increased over time to a greater extent in this study. The least squares (LS) mean change scores at Week 6 had standard deviations of 6.48 and 6.84, respectively, for gepirone and placebo, and the difference (95% CI) between treatments (-1.94; -3.84, -0.03) was statistically significant (p=0.046). By the Week 8 ET visit, the between-treatment difference was -1.5 points (-3.60, 0.59) and the standard deviations were 7.44 and 7.40 for gepirone ER and placebo groups, respectively (p=0.159). In fact, at Week 2, the between-group difference was comparable to that observed at the end of the study (-1.52 points), but with smaller standard deviations around the

means (4.55 and 4.28, respectively), this difference was statistically significant (p=0.016). These observations do nothing to change the study's weak results, but they do highlight the inherent variability of HAMD scores and the difficulty of reproducing drug effects from one study to the next in MDD.

For secondary efficacy variables, subjects treated with gepirone ER experienced numerically greater symptom improvement compared to placebo-treated subjects throughout the study, with occasional significant differences observed at Weeks 2 through 6. However, by Week 8, differences between treatments were not statistically significant at the 0.05 level.

Responder rates (percentage of subjects with 50% or greater reduction from baseline scores) for the HAMD-17, HAMD-25, HAMD-21, HAMD-28, CGI, and MADRS scales were consistently higher for the gepirone ER group compared to placebo. The HAMD-25 responder rate was significantly greater in the gepirone ER group at endpoint; however, responder rates for the other scales did not achieve significance at Week 8.

Likewise, the number of subjects classified as HAMD-17 remitters (subjects with post-baseline HAMD-17 score of ≤ 7) was numerically greater in the gepirone ER group than in the placebo group at all time points, but the differences failed to reach statistical significance.

Overall, this study failed to achieve statistical significance for the primary efficacy variable (change from baseline HAMD-17 at endpoint), but trends in mean values were evident at earlier visits for this and other secondary variables. Individual results of this study are not particularly strong, but taken together suggest that gepirone ER has a beneficial impact on symptoms of depression.

Active-Controlled Studies

Five studies included both placebo and active comparator arms, as shown in Table 11. Two studies, 134004 and 134006, were conducted in subjects with MDD with atypical features, referred to as Atypical Depression (AD). Three studies, CN105-052, CN105-053, and 134017, were conducted in subjects with MDD not further specified.

Table 11: Gepirone ER Active-Controlled Studies

Study Number	Indication	N/group	Active Control Drug	Primary Efficacy Variable
CN105-052	MDD	37	Fluoxetine	HAMD17 CFB
CN105-053	MDD	56	Imipramine	HAMD17 CFB
134004	AD	136	Fluoxetine	HAMD25 CFB
134006	AD	144	Paroxetine	HAMD25 CFB
134017	MDD	159	Fluoxetine	MADRS

The sponsor believes these studies lacked assay sensitivity because the active comparator failed to show superiority over placebo for the primary efficacy variables. While this rationale is appropriate as a basis for dropping studies from consideration (Laughren, TP. Eur Psych 16: 418-423, 2001), the ISE provided limited information to make this judgment. By contrast, the FDA

does not dismiss these studies from consideration and cites findings from secondary efficacy variables to conclude that the comparator did, in fact, demonstrate efficacy in some of these trials. As a result, the FDA designates several of the studies as "negative" trials because gepirone ER is not shown to be superior to placebo for the same secondary efficacy variables used to conclude assay sensitivity for the active comparator. The studies in question (CN105-052, CN105-053, 134004, 134006, and 134017) will be summarized individually below, ordered by date of completion.

Study CN105-052 [Jun 1991 – Aug 1992]

This was a 2-center, 8-week, randomized, double-blind, flexible dose study evaluating gepirone ER (20-60 mg/day), fluoxetine (10-40 mg/day), and placebo in subjects with non-psychotic MDD. For business reasons, the study was terminated early by BMS short of the planned 240 subjects. Only 111 subjects were randomized (36 gepirone ER, 37 fluoxetine, and 38 placebo). The original dose of gepirone ER (10-40 mg/day) was amended to 20-60 mg/day after 7 months; the average dose administered was 43.4±17.6 mg/day.

The primary analyses of HAMD-17 change from baseline and CGI responder rate did not detect statistically significant pair-wise differences between either gepirone ER and placebo or fluoxetine and placebo. No pair-wise gepirone ER-placebo differences were found in any of the secondary efficacy parameters.

Comment: Several factors are likely to have contributed to the failure of the study. First, the study was prematurely terminated by BMS with less than 40 subjects per treatment group, only 46% of the required sample size, thus reducing the power to detect treatment effects (under assumptions on p. 56 of the CSR) from 80% to 43%. Secondly, the placebo response rate was relatively high: 57% of placebo subjects were CGI responders at Week 8/Endpoint. This would tend to obscure evidence of response to active treatment. Finally, the dose of gepirone used in this study (average maximum dose of 43.4 mg/day) was relatively low within the proposed therapeutic dose range (40-80 mg/day). These factors and the inability to differentiate effects of the active control product (fluoxetine) given at therapeutic doses (10-40 mg) and placebo make this a failed study. Its failure to clearly differentiate treatment effects for both gepirone and the active comparator is most likely due to the fact that it was terminated early by BMS, which resulted in inadequate sample sizes and reduced power.

Study CN105-053 [Apr 1991 – Aug 1992]

This was a 2-center, randomized, double-blind, flexible dose study evaluating gepirone ER (10-60 mg/day), imipramine (50-200 mg/day) and placebo over an 8-week treatment period in subjects with non-psychotic MDD. The original dose of gepirone ER (10–40 mg/day) was amended to 10-60 mg/day after 5 months; the average dose administered was 50.4±13.9 mg/day.

The study was planned for 240 subjects, as this would provide 80% power to detect a difference of 4 points in HAMD-17 total score between each active drug and placebo at the Bonferroni-adjusted alpha level of 2.5% (assumed SD=8.1). However, BMS terminated the program after

170 subjects were randomized (58 gepirone, 56 imipramine, and 56 placebo), reducing the power to approximately 63%.

Overall, 46.4% of subjects (78/168 treated) discontinued the 8-week treatment period, primarily due to lack of efficacy (20.7% gepirone, 1.9% imipramine, and 50.0% placebo) and adverse events (12.1% gepirone, 27.8% imipramine, and 3.6% placebo). The mean maximum dose was 50.4 mg/day in the gepirone ER group and 169 mg/day in the imipramine group.

The primary endpoints were HAMD-17 change from baseline and CGI responder rate. For pooled study centers, no significant pairwise differences (gepirone ER-placebo or imipramine-placebo) were detected for the primary efficacy variables.

Table 12 summarizes endpoint results for all primary and secondary efficacy variables collected in this study. Trends favored each active treatment over placebo, but differences did not achieve statistical significance for change from baseline HAMD-17 at any time point except Week 2 (imipramine vs. placebo, p=0.041).

The sponsor notes that the 2 study sites showed substantially different efficacy results. Site 0001 (Feiger) completed enrollment of 121 subjects (101% enrolled) and showed statistically significant pair-wise differences between gepirone ER and placebo and between imipramine and placebo favoring active treatments for several efficacy endpoints (see Table 13 below). No significant pair-wise differences were detected in Site 0002 (Gelenberg, 47 subjects: 39% enrolled) for any of the primary or secondary efficacy parameters (Table 14).

Table 12: Efficacy Results in Study CN105-053

Efficacy Variable	End of Treatment Outcome (ITT/LOCF)*			Pairwise Tests (p-values)		
	Gepirone	Imipramine	Placebo	G vs. P	I vs. P	G vs. I
	N=56	N=54	N=56			
HAMD-25 CFB	-12.9	-14.5	-10.9	0.330	0.084	NR
HAMD-17 CFB	-9.7	-11.5	-9.0	0.687	0.144	NR
HAMD-28 CFB	-15.4	-17.5	-12.6	0.266	0.055	NR
HAMD-Item1 CFB	-1.0	-1.2	-0.9	0.929	0.200	NR
CGI (severity) CFB	-1.3	**-1.8**	**-1.1**	0.535	**0.021**	NR
CGI (global improvement)	2.4	**2.2**	**2.9**	0.110	**0.031**	NR
% Responders (HAMD-17)	43%	54%	38%	0.551	0.084	NR
% Responders (CGI)	54%	**72%**	**39%**	0.125	**<.001**	NR
MADRS CFB	-12.3	-15.4	-12.3	0.987	0.197	NR
HAMD Factor 1 CFB (anxiety/somatization)	-2.9	-3.3	-2.8	0.826	0.441	NR
HAMD Factor V CFB	-3.3	-3.9	-2.9	0.524	0.115	NR
HAMD Factor VI CFB	-1.8	-2.3	-1.8	0.945	0.216	NR
HAMA total score CFB	-7.0	-6.2	-6.7	0.859	0.705	NR

Source: CN105-053 Final Report Appendices F7.1, 7.2, and 7.4
LS means and p-values from ANOVA, with treatment and center as factors; CMH test for % responders
NR = Not Reported; CGI Responder = Much improved or very much improved
*For ITT, N=166 subjects were randomized, received treatment, and at least one post-baseline assessment.

Table 13: Efficacy Results in Study CN105-053 (SITE 001, Feiger)

Efficacy Variable	End of Treatment Outcome (ITT/LOCF)			Pairwise Tests (p-values)		
	Gepirone	Imipramine	Placebo	G vs. P	I vs. P	G vs. I
	N=41	N=39	N=40			
HAMD-25 CFB	-14.4	-14.6	-8.9	**0.007**	**0.006**	NR
HAMD-17 CFB	-10.1	-10.9	-6.8	**0.049**	**0.017**	NR
HAMD-28 CFB	-17.3	-17.6	-10.4	**0.006**	**0.005**	NR
HAMD-Item1 CFB	-1.2	-1.4	-0.6	**0.010**	**0.001**	NR
CGI (severity) CFB	-1.5	-1.7	-0.9	**0.031**	**0.004**	NR
CGI (global improvement)	2.4	2.1	3.1	**0.012**	**0.001**	NR
% Responders (HAMD-17)	44%	49%	33%	0.294	0.145	NR
% Responders (CGI)	56%	72%	33%	**0.034**	**0.001**	NR
MADRS CFB	-11.9	-14.3	-8.7	0.179	0.020	NR
HAMD Factor 1 CFB (anxiety/somatization)	-3.3	-3.4	-2.2	0.061	0.045	NR
HAMD Factor V CFB	-4.0	-4.0	-2.2	**0.011**	**0.009**	NR
HAMD Factor VI CFB	-1.0	-1.8	-1.3	0.610	0.205	NR
HAMA total score CFB	-7.5	-6.9	-5.2	0.106	0.218	NR

Source: CN105-053 Final Report Appendices F7.1, 7.2, and 7.4
LS means and p-values from ANOVA, with treatment as factor; CMH test for % responders
NR = Not Reported

Table 14: Efficacy Results in Study CN105-053 (SITE 002, Gelenberg)

Efficacy Variable	End of Treatment Outcome (ITT/LOCF)			Pairwise Tests (p-values)		
	Gepirone	Imipramine	Placebo	G vs. P	I vs. P	G vs. I
	N=15	N=15	N=16			
HAMD-25 CFB	-11.3	-14.3	-12.9	0.709	0.724	NR
HAMD-17 CFB	-9.3	-12.2	-11.1	0.589	0.776	NR
HAMD-28 CFB	-13.5	-17.3	-14.8	0.798	0.614	NR
HAMD-Item1 CFB	-0.7	-1.1	-1.3	0.163	0.550	NR
CGI (severity) CFB	-1.1	-1.8	-1.4	0.529	0.340	NR
CGI (global improvement)	2.5	2.4	2.7	0.718	0.638	NR
% Responders (HAMD-17)	40%	67%	50%	0.582	0.355	NR
% Responders (CGI)	47%	73%	56%	0.600	0.328	NR
MADRS CFB	-12.7	-16.5	-15.9	0.502	0.900	NR
HAMD Factor 1 CFB (anxiety/somatization)	-2.5	-3.1	-3.4	0.525	0.855	NR
HAMD Factor V CFB	-2.6	-3.8	-3.6	0.388	0.831	NR
HAMD Factor VI CFB	-2.6	-2.9	-2.3	0.750	0.539	NR
HAMA total score CFB	-6.5	-5.4	-8.2	0.499	0.294	NR

Source: CN105-053 Final Report Appendices F7.1, 7.2, and 7.4
LS means and p-values from ANOVA, with treatment as factor; CMH test for % responders
NR = Not Reported

Comments: Data from this study do not provide clear evidence that either gepirone ER or imipramine reduced symptoms of depression in MDD. Assay sensitivity is questionable, given that imipramine-placebo differences did not achieve statistical significance for both primary efficacy variables. FDA claims that the effect of imipramine on HAMD-17 was significant at endpoint (p=0.038), but I could not verify this. CGI variables (change from baseline severity, global improvement and % responders) showed significantly greater improvement in the imipramine group than in the placebo group. However, the lack of consistency in results of the 2 participating centers detracts from this finding.

Study CN105-053 was terminated prematurely, after only one of the 2 investigators completed enrollment. This may have contributed to their inconsistent results. The test for treatment-by-center interaction was not statistically significant for change from baseline HAMD-17 (p=0.317; see Table 7.1.1-3, Appendix F). However, a test for interaction has limited power in this study, given the relatively small and unequal numbers of subjects enrolled in each center. The interpretation of the study is especially problematic because it has assay sensitivity for one of the primary efficacy variables (CGI), but not both. The largest study site (Feiger) demonstrates highly significant treatment effects for both gepirone ER and the active comparator. However, results in the combined centers are inconclusive given the conflicting results of individual investigators, and the lack of clear assay sensitivity.

Study 134004 [Jun 2000 – Jul 2002]

This was a 10-center, 3-arm, randomized, double-blind study evaluating gepirone ER in comparison to placebo and fluoxetine over an 8-week treatment period in subjects diagnosed with MDD having "atypical features." The design allowed flexible gepirone ER doses (20-80 mg/day with a starting dose of 20 mg/day and forced titration to 40 mg/day during the first week; further titration to between 40 and 80 mg/day was allowed) and flexible fluoxetine dose (20 mg/day for the first 4 weeks with titration allowed up to 40 mg/day afterwards). The mean final dose of gepirone was 67.1 mg/day (±19.2), with 79.3% of subjects titrated to 60-80 mg/day. The mean final dose of fluoxetine was 34.1 mg/day (±9.2), with 70.3% at the top dose.

Overall, 409 subjects were randomized and treated (135 gepirone ER, 138 fluoxetine, and 136 placebo). The drop-out rate was higher in the gepirone ER group (36.3%) than in the fluoxetine (18.1%) or placebo (21.3%) groups, mainly due to adverse events (10.4% vs. 2.9% and 1.5%) and "other reasons" (22.2% vs. 12.3% and 16.9%) including lost to follow-up, non-compliance, and withdrawn consent. A total of 391 subjects with post-baseline data comprised the ITT population (125 gepirone ER, 136 fluoxetine, and 130 placebo).

No statistically significant pair-wise differences were detected between treatment groups for the primary efficacy variable (change from baseline HAMD-25 total score) or for any of the pre-defined secondary efficacy variables: change from baseline scores for HAMD-17, HAMD-28, HAMD-item 1 (depressed mood), and global impression of severity (CGI item 1), global impression of change (CGI item 2), HAMD-25 responders (decrease from baseline ≥ 50%), CGI responders ("much" or "very much improved" on the CGI change score), and HAMD-25

remitters (both HAMD-17 total score ≤ 7 and HAMD-25 total score ≤ 10 on a post-baseline assessment).

A summary of results for all efficacy variables is presented in Table 15.

Table 15: Efficacy Results in Study 134004

Efficacy Variable	End of Treatment Outcome (ITT/LOCF)			Pairwise Tests (p-values)		
	Gepirone	Fluoxetine	Placebo	G vs. P	F vs. P	G vs. F
HAMD-25 CFB	-9.76	-11.66	-10.63	0.416	0.325	0.089
HAMD-17 CFB	-5.67	-7.5	-6.55	0.282†	NR	0.027
HAMD-28 CFB	-11.54	-14.0	-12.52	0.438	NR	NR
HAMD-Item1 CFB	-0.97	-1.2	-1.11	0.328	NR	NR
CGI (severity) CFB	-0.98	-1.2	-1.11	0.392	NR	NR
CGI (global improvement)	2.98	2.7	2.76	0.142	NR	NR
% Responders (HAMD-25)	33.87%	NR	36.15%	0.765	NR	NR
% Responders (CGI)	34.68%	NR	42.42%	0.224	NR	NR
% Remitters (HAMD-25)	16.94%	NR	23.85%	0.178	NR	NR
HAMD-31 (items 22-26) CFB	-2.82	-2.57	-2.80	0.948	NR	0.459
HAMD Factor 1 CFB (anxiety/somatization)	-2.09	-2.56	-1.93	0.594	0.154	NR
HAMA total score CFB	-4.08	-5.68	-4.95	0.226	NR	0.025

LS means and p-values from ANOVA, with treatment and center as factors; CMH test for % responders.
†Significant treatment-by-center interaction (p=0.05) for HAMD-17 CFB.
NR = Not Reported

The sponsor concludes: Failure of fluoxetine to demonstrate efficacy on the primary parameter, when compared to placebo, indicates that this was a failed study.

Comment: Based on the sponsor's analysis, the active comparator (fluoxetine) had no significant effect on the primary efficacy variable (change from baseline HAMD-25). FDA cites a significant difference favoring fluoxetine over gepirone ER for change from baseline HAMD-17 score at end of study (p=0.027); this p-value was not reported in the NDA. The FDA may have applied an ANOVA model using data from all treatment groups to perform all pair-wise comparisons, whereas the sponsor applied separate ANOVA models to each drug-placebo pair for all but the primary variable. Note that the sponsor's comparison of gepirone and placebo groups showed evidence of treatment-by-center interaction for HAMD-17 (p=0.05; page 205 Appendix F), indicating that gepirone's effect on this variable was not consistent across centers. Center-specific results were not reported in the NDA, but it would be useful to determine if the interaction is qualitative and, if so, how many of the 10 centers favor gepirone over placebo (or vice versa) for HAMD-17. It would also be important to determine how such interaction might have affected the FDA's analysis of HAMD-17.

Nevertheless, even if we ignore the interaction issue, this single significant difference favoring fluoxetine over gepirone is not a very convincing measure of assay sensitivity because (1) it is not supported by the primary efficacy variable or by other secondary efficacy variables at the

same or earlier time points, and (2) it might be a spurious finding, given the large number of pairwise significance tests performed.

The protocol did not require any minimum level of symptom severity, so subjects enrolled in this study had a low degree of depression. They were selected for symptoms of atypical depression, a sub-type of MDD. In fact, SSRIs are not known to be effective (or labeled for use) in this specific population, which might explain the lack of discernible drug effects. The average HAMD-17 score at baseline was relatively low (19.6) compared to values of 23-24 reported for studies 134001 and FK-GBE-007. Likewise, over half of subjects in the study scored 30 or less on the Bech Depression Inventory II (BDI II) scale. It is possible that the study failed to demonstrate drug-placebo differences because subjects were enrolled without sufficient symptoms, providing little opportunity to show improvement.

In my opinion, the study does not provide a reliable basis for judging the efficacy of gepirone because the active comparator drug (fluoxetine) failed to demonstrate a significant effect on the primary efficacy variable. Results based on other outcome variables were not sufficient to infer that fluoxetine was effective in this study. Therefore, this study does not have adequate assay sensitivity to gauge the efficacy of gepirone ER.

Study 134006 [Dec 2000 – Oct 2003]

This was a 13-center, 3-arm, randomized, double-blind study evaluating gepirone ER in comparison to placebo and paroxetine over an 8-week treatment period in subjects with atypical depression. Similar in design to study 134004, this study enrolled subjects with HAMD-25 total scores ≥ 18 at baseline, and the change from baseline in HAMD-25 was used as the primary efficacy endpoint. The design allowed flexible gepirone ER doses (20-80 mg/day with a starting dose of 20 mg/day and forced titration to 40 mg/day during the first week; further titration to between 40 and 80 mg/day was allowed) and flexible paroxetine doses (10-40 mg/day with starting dose of 10 mg/day and forced titration to 20 mg/day during the first week; further titration to between 20 and 40 mg/day was allowed). The mean final dose of gepirone was 67.3 mg/day (±17.4), with 82.3% of subjects titrated to 60-80 mg/day. The mean final dose of paroxetine was 33.9 mg/day (±8.9), with 62.7% at the top dose.

A total of 437 subjects were treated (147 gepirone ER, 142 paroxetine, and 148 placebo). The drop-out rate was higher in the gepirone ER group (31.3%) than in the paroxetine (28.9%) or placebo (24.3%) groups, mainly due to adverse events (12.2% vs. 5.6% and 2.7%) and lack of efficacy (6.1% vs. 2.8% and 4.7%). Most drop-outs were attributed to "other reasons" (primarily lost to follow-up and withdrew consent): 12.9% gepirone ER, 20.4% paroxetine, and 16.9% placebo. A total of 422 subjects with post-baseline data comprised the ITT population (143 gepirone ER, 136 paroxetine, and 143 placebo).

No significant treatment effects were detected for either gepirone ER or paroxetine based on the primary efficacy variable (change from baseline HAMD-25 total score) or any of the pre-defined secondary efficacy variables: change from baseline scores for HAMD-17, HAMD-28, HAMD-item 1 (depressed mood), and global impression of severity (CGI item 1), global impression of change (CGI item 2), HAMD-25 responders (decrease from baseline ≥ 50%), CGI responders

("much" or "very much improved" on the CGI change score), and HAMD-25 remitters (both HAMD-17 total score ≤ 7 and HAMD-25 total score ≤ 10 on a post-baseline assessment).

A summary of results for all efficacy variables is presented in Table 16.

Table 16: Efficacy Results in Study 134006

Efficacy Variable	End of Treatment Outcome (ITT/LOCF)			Pairwise Tests (p-values)*		
	Gepirone	Paroxetine	Placebo	G vs. P	Px vs. P	G vs. Px
HAMD-25 CFB	-10.94	-12.58	-11.00	0.953	0.178	0.209
HAMD-25 CFB (per ISE)†	-10.93	-12.58	-11.03	0.993†	0.060†	NR
HAMD-17 CFB	-6.92	-9.1	-7.15	0.750	0.026‡	0.042‡
HAMD-28 CFB	-12.68	-14.8	-12.57	0.927	NR	NR
HAMD-Item1 CFB	-1.11	-1.4	-1.07	0.720	NR	NR
CGI (severity) CFB	-1.10	-1.4	-1.21	0.423	NR	NR
CGI (global improvement)	2.68	2.3	2.63	0.726	NR	NR
HAMD Bech-6 score CFB	-4.60	-6.0	-4.55	0.918	NR	NR
% Responders (HAMD-25)	42.86%	NR	41.96%	0.894	NR	NR
% Responders (CGI)	45.71%	NR	46.85%	0.808	NR	NR
% Remitters (HAMD-25)	30.71%	NR	32.87%	0.652	NR	NR
HAMD-31 (items 22-26) CFB	-2.76	-2.10	-2.47	0.253	NR	0.012
HAMD Factor 1 (anxiety)	-2.26	-2.9	-2.25	0.989	NR	NR

Source: 134006 Final Report Appendix F Tables 6.1.1.4AA-6.1.2.3.3
*P-values and LS means based on ANCOVA model applied to 2 groups (gepirone vs. placebo), with treatment and center as factors, baseline value as covariate.
LS means and p-values not reported for paroxetine vs. placebo.
NR=Not Reported
†Pairwise tests of each treatment vs. placebo were reported in the ISE (page 553) for the primary efficacy variable (HAMD-25 CFB), but no other variables.
‡P-values for HAMD-17 were obtained from FDA correspondence.

Significant differences were detected between the effects of gepirone ER and paroxetine on HAMD hypersomnia/hyperphagia factor (items 22-26) at Visit 3 and Visit 6/ET using both the LOCF and OC approaches. It appeared that gepirone reduced hypersomnia/hyperphagia more than paroxetine, but neither drug's effect on this factor was significant compared to placebo.

The sponsor concludes: Failure of paroxetine to demonstrate efficacy on the primary parameter, when compared to placebo, indicates that this was a failed study.

Comment: Neither gepirone ER nor the comparator drug (paroxetine) demonstrated a statistically significant effect on the primary efficacy variable (change from baseline in HAMD-25 total score). Lacking assay sensitivity, this study has questionable value as a basis for judging the efficacy of gepirone ER. However, the FDA highlights significant differences in HAMD-17 total scores favoring paroxetine over placebo (p=0.026) and favoring paroxetine over gepirone (p=0.042) at end of study. These p-values were not reported in the NDA and might have been based on a re-analysis performed by FDA. I suspect FDA used a model that included data from all treatment groups to perform all pair-wise comparisons, whereas the sponsor's analysis is

confined to 2-group drug-placebo comparisons. In either case, the HAMD-17 findings are not supported by the primary efficacy variable, so they provide rather weak evidence of assay sensitivity without more clear-cut trends or consistent findings for other efficacy variables. Even if we accept the difference in HAMD-17 scores as evidence that paroxetine was efficacious, its effect on this score fails to validate treatment comparisons based on HAMD-25, the scale used to select eligible subjects (total score ≥ 18) and assess symptoms most relevant to subjects with atypical depression.

Based on the sponsor's ANCOVA analysis of HAMD-25 (paroxetine vs. placebo), there is evidence of treatment-by-center interaction (p=0.024; p. 248 Appendix F), suggesting that paroxetine's effect on this measure was not consistent across centers. A similar test of interaction was not performed for HAMD-17. It would be worth examining center-specific results, to see how many of the 13 centers showed trends favoring fluoxetine over placebo for HAMD-17 and HAMD-25. This may also affect FDA's assessment of assay sensitivity. Without clear evidence that paroxetine had a significant effect on the primary efficacy variable, it is difficult to draw firm conclusions from this study.

The failure of this study might be attributable to the low severity of depression symptoms in the subjects enrolled. The sponsor notes that subjects with atypical depression tend to have very low baseline HAMD-17 scores. In this study, the average baseline score was 19.0, compared to 23 or 24 for Studies 134001 and FK-GBE-007. Nearly half of subjects had low BDI II scores at baseline: 16% below 20 and 48% below 30 points. In subjects with values below 30, the final percent responders (reduction in BDI II ≥ 50%) were: placebo 47%, gepirone ER 37%, and paroxetine 52%; in subjects with BDI II values above 30 at entry, response rates were: placebo 37%, gepirone ER 60%, and paroxetine 56%. In the subgroup with more severe symptoms, response rates were significantly higher for both gepirone ER (p=0.009) and paroxetine (p=0.02) compared to placebo. These results suggest that gepirone ER and paroxetine provided benefit in the subgroup of subjects with more severe depression.

Study134017 [Oct 2002 – Jan 2004]

This was a 9-center, randomized, double-blind study evaluating gepirone ER (20-80 mg/day) in comparison to placebo and fluoxetine over an 8-week treatment period in 496 subjects with MDD (165 gepirone ER, 166 fluoxetine, and 165 placebo). The design allowed flexible doses of gepirone ER (20-40 mg/day during the first week and 40-80 mg/day thereafter) and fluoxetine (20 mg/day for the first 4 weeks and 20-40 mg/day thereafter). Of the 496 subjects randomized, 16 were excluded from the ITT dataset (1 was untreated and 15 had no post-baseline assessments).

The average dose for gepirone ER was 58.7±15.0 mg/day, with 62.5% of subjects titrated to 80 mg/day. The average dose of fluoxetine was 25.9±4.7 mg/day, with 72.7% at the top dose. The drop-out rate was higher in the gepirone ER group (31.5%) than in the fluoxetine (24.1%) or placebo (21.3%) groups, mainly due to adverse events (8.5% vs. 4.8% and 1.2%) and "other reasons" (17.6% vs. 16.9% and 15.9%) including lost to follow-up, non-compliance, and

withdrawn consent. Twenty subjects dropped out due to lack of efficacy (5.5% gepirone ER, 2.4% fluoxetine, and 4.3% placebo).

The primary efficacy variable was change from baseline in MADRS total score. No statistically significant pairwise differences were detected between gepirone ER and placebo or between fluoxetine and placebo for the primary endpoint. For other variables, pairwise tests of treatment differences were performed only if the overall treatment effect was statistically significant at the .05 level. A summary of results for all efficacy variables is presented in Table 17.

Table 17: Efficacy Results in Study 134017

Efficacy Variable	End of Treatment Outcome (ITT/LOCF)			Pairwise Tests (p-values)*		
	Gepirone	Fluoxetine	Placebo	G vs. P	F vs. P	G vs. F
	N=160	N=161	N=159			
MADRS score CFB	-12.23	-13.88	-12.73	0.650	0.299	0.136
HAMD-25 CFB	-12.03	-13.77	-12.86	--	--	--
HAMD-17 CFB	-10.23	-11.76	-10.96	--	--	0.042†
HAMD-Item1 CFB	-1.33	-1.50	-1.24	--	--	--
CGI (severity) CFB	-1.24	-1.41	-1.32	--	--	--
CGI (global improvement)	2.48	2.29	2.44	--	--	--
% Responders (HAMD-17)	42.14%	57.23%	45.28%	0.607	--	0.006
% Responders (CGI)	54.09%	62.26%	52.83%	0.829	--	0.152
% Remitters (HAMD-17)	22.01%	32.08%	31.45%	0.060		0.044
HAMD-31 (items 22-24) CFB Hypersomnia factor	-0.59	-0.43	-0.65	--	--	--
HAMD-31 (items 23 and 26) Hyperphagia factor	-0.10	-0.38	-0.06	0.742	0.016	0.037

Source: 134017 Final Report Appendix F
*Based on ANOVA model, with treatment as the main effect; CMH test for % responders.
†Reported by FDA. A similar result can be derived from Table 6.2.1-7.B (page 334 Appendix F). The overall test of treatment effect was not significant (p=0.137), but the pairwise test of the difference between gepirone and fluoxetine yields Z=1.5343/0.7683=1.997, with p=0.046.

For change from baseline in HAMD-17, the treatment effect was not statistically significant at end of study (p=0.137) or at any other time point except Visit 1 (p=0.025). At Visit 1, the reduction in HAMD-17 was greater in the fluoxetine group compared to either the placebo (p=0.01) or gepirone (p=0.045) groups. No significant treatment effects were noted for change from baseline in HAMD-25, HAMD item 1 score, or CGI severity score, the CGI change item, or the HAMD-17 responders. The change from baseline in the HAMD hypophagia factor was the only secondary efficacy variable showing significant differences between gepirone and fluoxetine, with greater reductions on fluoxetine at all visits including endpoint.

Comment: The effect of gepirone was indistinguishable from placebo for all efficacy variables in this study. The active control (fluoxetine) also failed to show a significant effect beyond that of placebo for the primary efficacy variable. While a few secondary efficacy variables showed directional trends favoring fluoxetine over placebo, these small differences were not sustained

over the course of the treatment period and achieved statistical significance only rarely at sporadic time points. These marginal effects of the active control on secondary variables are not sufficient to validate the assay sensitivity of the trial. If fluoxetine had demonstrated a clear and convincing advantage over placebo, the lack of significant findings for gepirone could be interpreted as true lack of efficacy. Instead, the study provides no evidence that either drug performed significantly better than placebo.

The failure of this study to detect treatment effects may be attributable to the placebo response rate, which was higher than expected. This may have been exacerbated by enrollment of subjects without severe enough levels of depression. As noted by the sponsor, subjects were selected for this study based on HAMD-17 total scores (HAMD-17 $\geq$ 18) rather than the primary efficacy variable (MADRS). This raises the concern that subjects were enrolled without severe enough levels of depression on the primary scale. The sponsor cites greater placebo response rates in subjects with mild illness (MADRS scores < 30) than in those with more severe symptoms. These observations may account for the lack of discernable treatment effects in this study.

General comment related to active-controlled studies: In my opinion, the FDA used rather liberal criteria for accepting assay sensitivity of the active controlled studies (e.g., single instances of statistically significant drug-placebo differences in studies 134004, 134006, and 134017 -- see FDA meeting minutes dated Jan 8, 2008) to deem the studies valid for judging the efficacy of gepirone ER. However, for these studies to be interpretable, the effect of the active comparator should be more clearly demonstrated, at least on the basis of the primary efficacy variable used to design and size the study.

Meta - Analyses

The ISE document included a meta-analysis to estimate the overall effect of gepirone ER compared to placebo based on all available evidence from controlled clinical studies evaluating its short-term efficacy. The sponsor performed separate meta-analyses by combining all 12 controlled clinical studies in the NDA, as well as subsets of the studies that were deemed pivotal (2 studies) or supportive (5 studies) of gepirone's efficacy. A treatment difference (gepirone ER vs. placebo) was considered statistically significant if the p-value was below 0.05.

Table 18 summarizes results of the meta-analyses based on HAMD-17 (change from baseline to endpoint of double-blind treatment period):

Table 18: Meta-Analysis: Change in HAMD-17 from Baseline to End of Double-Blind Treatment Period in All Clinical Studies of Gepirone ER in MDD (12 Studies)

Study (1)	Number of subjects (N)		(Adjusted) mean change		Treatment difference and 95% CI		p-value
	Gepirone ER	Placebo	Gepirone ER	Placebo	(Gepirone ER - Placebo)	SE	
Pivotal Studies							
ORG 134001	101	101	-9.04	-6.57	-2.47 (-4.41, -0.53)	0.98	0.013
FKGBE007	116	122	-10.24	-7.79	-2.45 (-4.47, -0.43)	1.02	0.018
Supportive Studies							
ORG 134023 (2)	123	123	-7.93	-8.05	0.13 (-1.79, 2.04)	0.97	0.898
FKGBE008	96	99	-9.86	-8.48	-1.38 (-3.48, 0.71)	1.06	0.195
ORG 134002	102	103	-9.95	-9.24	-0.71 (-2.44, 1.02)	0.88	0.417
CN105-078	88	47	-7.42	-6.42	-1.00 (-3.17, 1.16)	1.10	0.362
CN105-083	73	39	-9.46	-8.97	-0.49 (-3.52, 2.53)	1.53	0.747
Studies Lacking AS							
ORG 134017	159	159	-10.36	-10.99	0.63 (-0.88, 2.13)	0.76	0.412
ORG 134004	124	130	-5.63	-6.66	1.03 (-0.55, 2.61)	0.80	0.199
CN105-052	35	37	-10.94	-10.28	-0.66 (-4.86, 3.55)	2.11	0.757
ORG 134006	140	143	-6.89	-7.13	0.24 (-1.18, 1.66)	0.72	0.742
CN105-053	56	56	-10.20	-8.15	-2.05 (-4.96, 0.87)	1.47	0.167
Meta-analyses (3)							
Supportive/Pivotal Studies Combined	699	634			-1.22 (-1.99, -0.45)	0.39	0.002
Treatment by study interaction							0.470
Supportive Studies Combined	482	411			-0.68 (-1.60, 0.24)	0.47	0.149
Treatment by study interaction							0.874
All 12 Studies Combined	1213	1159			-0.48 (-1.03, 0.08)	0.28	0.093
Treatment by study interaction							0.108

(1) Study statistics obtained using ANCOVA model with terms for treatment and center, and baseline value as covariate.
(2) Considered a negative study, but included with supportive studies for the purpose of meta-analysis.
(3) Combined estimates of the gepirone-placebo difference obtained as weighted averages of the gepirone-placebo differences with reciprocals of the squares of the standard errors of the by-study differences used as the weights. The standard errors of the overall estimates are the reciprocals of the square roots of the sums of the weights.
Source: Table 4 of the ISE.

Meta-analyses in these study groupings were also carried out for other efficacy variables (CGI responders, MADRS, Bech-6, HAMD-14, HAMD-Item 1, and 3 sleep-related items from HAMD. Pooled results for the 12 clinical studies are shown in Table 19.

Table 19: Meta-Analyses – Pooled Results from All 12 Clinical Studies (ITT/LOCF)

Efficacy Variable	Treatment Difference (Gepirone – Placebo)	95% Confidence Interval	p-values	
			Treatment Effect	Treatment by Study Interaction
HAMD-17 CFB	-0.48	(-1.03, 0.08)	0.093	0.108
CGI Response (odds ratio)	1.25	(1.04, 1.49)	0.015	0.422
MADRS CFB	-1.71	(-2.65, -0.77)	< 0.001	0.361
Bech-6	-0.53	(-0.85, -0.20)	0.002	0.171
HAMD-Item1 CFB	-0.14	(-0.22, -0.05)	0.002	0.168
HAMD-14 CFB	-0.63	(-1.11, -0.15)	0.010	0.210
HAMD (sleep items) CFB	0.12	(-0.02, 0.26)	0.085	0.008

Source: Tables 4-8, 8.1 and 8.2 of ISE

Comment: Pooling 12 studies provides a comprehensive picture of the treatment effect across all randomized, controlled studies in the clinical development program, without selection based on quality or outcome. The pooled result for HAMD-17 did not reach statistical significance, with some evidence of inconsistency among studies (p=0.108 for heterogeneity). Tests of heterogeneity (treatment-by-study interaction) were not statistically significant for any of the other variables, which help to justify the pooling strategy. Overall, the studies showed significant treatment effects for the odds of CGI response, and change from baseline for MADRS, Bech-6, HAMD-Item 1 (depressed mood), and HAMD-14. Despite variation in study outcomes and deficiencies related to assay sensitivity in the active control trials, the meta-analysis demonstrated statistically positive results for gepirone ER compared to placebo for these efficacy variables.

The FDA also performed a meta-analysis of HAMD-17 results from 10 gepirone ER studies (non-approvable letter dated November 2, 2007), including all but two clinical studies (FKGBE007 and 134001) that were positive in demonstrating the efficacy of gepirone. The FDA excluded these studies "to determinate if, among the remaining 10 non-supportive trials, there was any suggestion of an effect for gepirone ER." This rationale for excluding positive trials ignores a key goal of meta-analysis: to assess the overall treatment effect based on combined evidence from all available randomized controlled trials. Criteria for study exclusions should be restricted to deficiencies in randomization, blinding, protocol compliance, or data integrity that would introduce bias in efficacy outcomes.

Summary and Conclusions

The 12 studies submitted to FDA in support of marketing approval for gepirone ER in MDD represent over 20 years of clinical development by three different sponsors. Several of the studies failed to meet their primary efficacy objective and/or failed to show efficacy of the active comparator antidepressant (fluoxetine, paroxetine, or imipramine). Reasons for failure are open to conjecture, but might include: early termination of BMS studies (reduced power), inappropriate subject selection criteria (inclusion of subjects with atypical depression or low depression severity in studies conducted by Organon), heterogeneous subjects, lack of calibration or inconsistency in assessment scales, excessively rapid enrollment of subjects by multiple investigators, and/or the use of gepirone ER doses below the appropriate therapeutic range. Some of the studies were terminated early, before sufficient numbers of subjects were recruited, while others completed enrolment rapidly at multiple sites in a short period. All of these factors would lead to excess variability in efficacy outcomes.

These studies cannot all be weighted equally in terms of their quality or strength of evidence for efficacy conclusions. Table 20 provides a brief overview of the study designs and outcomes.

Based on my review, the studies can be classified as follows:

Five studies appear to be adequate and well-controlled for the purpose of judging the short-term efficacy of gepirone ER:

- Positive studies:
    - Studies **134001** and **FKGBE007** achieved statistical significance on the primary endpoint, and demonstrate the efficacy of gepirone ER for the treatment of MDD. FDA has also agreed that these 2 studies had positive findings.
- Supportive studies:
    - Study **134002** failed to achieve significance on the primary endpoint (HAMD-17 at final visit) but provided limited evidence of efficacy based on secondary efficacy variables and 'post-hoc' mixed-models analyses.
    - Study **FKGBE008** failed to achieve significance on the primary endpoint (HAMD-17 at final visit), but trends were evident at earlier visits for this and other secondary efficacy variables.
- Negative study:
    - Study **134023** was adequately designed with sufficient power, but offered no evidence that gepirone improved depression symptoms.

Seven studies were inconclusive regarding the efficacy of gepirone, because of specific design flaws and issues that limit their interpretability:

- Comparator failed / lack of assay sensitivity:

    - Study **CN105-053** was terminated early, with fewer than 60 subjects per treatment group, and the 2 participating study sites showed contradictory results. Data from the largest site showed positive effects for both gepirone and the active control (imipramine), whereas overall results in the combined sites were inconclusive due to the lack of assay sensitivity for the primary efficacy parameter, HAMD-17 change from baseline.

    - In study **134006**, the comparator drug (paroxetine) failed to demonstrate a statistically significant effect on the primary efficacy variable, HAMD-25 CFB. Significant differences in HAMD-17 favored paroxetine over both placebo and gepirone at endpoint, but these results are not sufficient to confirm the efficacy of the active control or validate treatment comparisons for gepirone ER based on the primary efficacy variable (HAMD-25), the scale used to select subjects for the trial. Further, the sub-type of MDD under study (atypical depression) differs from the population enrolled in the other trials.

    - Study **134004** failed to demonstrate that the active comparator (fluoxetine) had a significant effect on the primary efficacy variable (HAMD-25). Its effects on secondary variables were not sufficient to validate its use as a basis for judging the efficacy of gepirone ER. Subjects were enrolled with atypical depression, rather than MDD, which might explain the lack of assay sensitivity.

    - In study **134017**, the active comparator (fluoxetine) failed to demonstrate a statistically significant effect on the primary efficacy variable (MADRS). Significant differences favored fluoxetine over placebo for a few secondary variables at specific time points, but the findings were not strong or consistent enough to confirm the effects of fluoxetine and validate the assay sensitivity of the trial to the effects of gepirone ER.

- Failed study / low power:

    - Study **CN105-052** enrolled less than half the required sample size, thus reducing the power to detect treatment effects (43% power). The active control (fluoxetine) was indistinguishable from placebo based on all efficacy variables. Moreover, the dose of gepirone used in this study (average 43.4 mg/day) was low relative to its therapeutic dose range (60-80 mg/day).

    - Study **CN105-078** was planned with low power, terminated early (62% power), and failed its primary endpoint (HAMD-17 in pooled dose groups). The study employed inadequate doses of gepirone ER in half of the subjects; findings in the high dose group were positive, but not a reliable basis to generalize efficacy conclusions.

    - Study **CN105-083** was terminated early with insufficient power (53%) to detect drug-placebo differences and inconsistent results by center, so results were inconclusive.

Despite concerns about assay sensitivity and other design issues noted above, meta-analyses showed statistically significant results for gepirone ER compared to placebo based on several key efficacy variables (CGI response, MADRS, Bech-6, HAMD-Item 1, and HAMD-14), lending supportive evidence to the efficacy of gepirone ER in MDD.

Finally, it should be noted that this statistical review evaluated only the efficacy data from the controlled studies, without any critical assessment of short- or long-term safety data. As a result, no conclusions can be drawn with respect to risk-benefit determinations for this product.

Table 20: Summary Overview: Controlled Clinical Studies of Gepirone ER

Study # (Disease)	N/group Planned	N/group Actual	Control	Primary Efficacy Variable	Comparison	p-values*	Dose range / avg† [mg/day]	Comments
Positive								
134001 (MDD)	100	101 5 centers 8 weeks	Placebo	HAMD17 CFB	G vs. P -9.04 vs. -6.57	0.018	20-80 /61.1	Significant treatment effects; consistent trends in HAMD17 for 4 of 5 centers; center 1 (p=.011)
FKGBE007 (MDD)	120	115 9 centers 8 weeks	Placebo	HAMD17 CFB	G vs. P -10.24 vs. -7.79	0.016	20-80 / 58.2	Significant treatment effects; consistent trends in HAMD17 for 5 of 8 centers; 2 centers significant; Treatment-by-center interaction p=0.092
Supportive								
134002 (MDD)	100	107 5 centers 8 weeks	Placebo	HAMD17 CFB	G vs. P	0.417	20-80 / 57.9	Not significant for primary, positive trends; significant treatment effect using proc mixed for secondary variables.
FKGBE008 (MDD)	100	98 8 centers 8 weeks	Placebo	HAMD17 CFB	G vs. P	0.195	20-80 / 60.0	Not significant for primary, only trends and spotty significant in secondary variables.
Negative								
134023 (MDD)	120	123 12 centers 8 weeks	Placebo	HAMD17 CFB	G vs. P	0.898	20-80 / 61.3	No trends or significance for any variables.
Comparator Failed / Lack of Assay Sensitivity								
CN105-053 (MDD)	60	56 2 centers 8 weeks	Imip	HAMD17 CFB	G vs. P	0.687	10-60 / 46.4	Stopped early, 63% power; largest center shows positive effects for both Gep and Imip.
					I vs. P	0.144 (FDA: 0.038)		Imip>P for CGI, not HAMD17
					G vs. I	0.362		
134006 (AD)	142	147 13 centers 8 weeks	Parox	HAMD25 CFB	G vs. P	0.928	20-80 / 55.3	No significant drug-placebo differences for primary (HAMD25). No trends or significance for G vs. P; Treatment-by-center interaction (p=.024) for HAMD25.
					Prx vs. Pbo	0.178 (FDA: 0.026)		Comparator failed on primary variable.
					G vs. Prx	0.209 (FDA: 0.042)		Prx > Gep for HAMD17, per FDA, not primary var; Treatment-by-center interaction not investigated.

Study # (Disease)	N/group Planned	N/group Actual	Control	Primary Efficacy Variable	Comparison	p-values*	Dose range / avg† [mg/day]	Comments
134004 (AD)	130	125 10 centers 8 weeks	Fluox	HAMD25 CFB	G vs. P	0.416	20-80 / 58.7	No significant drug-placebo differences or trends for primary (HAMD25) or any other variables.
					F vs. P	0.325		Comparator failed on primary variable.
					G vs. F	0.089 (FDA: 0.027)		Possibly spurious finding
134017 (MDD)	150	160 9 centers 8 weeks	Fluox	MADRS	G vs. P	0.650	20-80 / 58.7	No positive trends or significant effects of G vs P for any variable
					F vs. P	0.289		Comparator failed on primary variable.
					G vs. F	0.136 (FDA: 0.042)		F > Gep for HAMD17, per FDA; not primary variable (MADRS)
Failed Study / Low Power								
CN105-052 (MDD)	60	35 2 centers 8 weeks	Fluox	HAMD17 CFB	G vs. P	0.825	10-60 / 38.7	Stopped early, 43% power No significant treatment effects; failed study
					F vs. P	0.798		Comparator failed on primary variable.
					G vs. F	0.916		
CN105-078 (MDD)	60	48/40 2 centers 6 weeks	Placebo	HAMD17 CFB	G vs. P	0.262	low: 10-50/30.4 high: 20-100/52.6	Terminated early; 62% power. No significant effect for pooled doses, trends for high dose.
CN105-083 (MDD)	60	36/37 2 centers 6 weeks	Placebo	HAMD17 CFB	G vs. P	0.747	low: 10-50/30.4 high: 20-100/57.1	Terminated early; 53% power. No significant effect for pooled doses; trends for high dose. Treatment-by-center interaction (p=.098); two centers had opposite results.

* FDA p-value based on HAMD17 Total Score, not always the primary efficacy parameter
All other data from 2007 NDA amendment (Intent-to-Treat population)
†Overall mean daily dose of gepirone ER during study

Diseases: MDD=Major Depressive Disorder, AD=Atypical Depression
CFB = change from baseline